Министерства высшего и среднего специального образования
Республики Узбекистан
“Ташкентский государственный технический университет имени Ислама Каримова”

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Кузбасский государственный технический университет имени Т. Ф. Горбачёва»

Е. А. Николаева, Ш.Т. Пирматов, Е. Н. Грибанов, А. Абдукаримов

МАТЕМАТИКА ДЛЯ ЭКОНОМИСТОВ. ЧАСТЬ 3

Теория вероятностей и математическая статистика

Учебное пособие

Тошкент-2019

Математика для экономистов.Часть-3. Теория вероятностей и математическая статистика / Е. А. Николаева, Ш.Т. Пирматов, Е. Н. Грибанов, А. Абдукаримов. Ташкент: ТГТУ, 2019г. 220 с.

Учебное пособие составлено в соответствии программой дисциплин «Математика для экономистов» приведены теории и основные теоремы по разделу теории вероятностей и математическое статистике. Изложен большое количество разобранных примеров и задач как по теории вероятностей так и математической статистике.

В приложении данного пособия приведено большое количество справочного материала, которой может быть востребовано при самостоятельном решении задач по теории вероятности и математической статистики.

Учебное пособие предназначена для студентов экономических направлений, оно будет также полезно для студентов, обучающихся по смежным направлениям.

Рецензенты:

Каган Е.С.	-	Заведующий кафедрой «Прикладная математика» КемГУ, к. т. н., доцент;
Худоеров Б.А.	-	Заведующий кафедрой «Высшая математика» ТИИИМСХ, д.т.н., профессор;
Гутова С.Г.	-	Доцент кафедры «Прикладная математика» КемГУ, к.т.н., доцент;
Эсонов Э.Э.	-	Доцент кафедры «Высшая математика» ТГТУ, к. ф.-м.н., доцент.

Учебное пособие рассмотрено и рекомендовано к печати научно – методическим советом ТГТУ им Ислама Каримова (протокол № ____ от ____ 2019 г.)

ПРЕДИСЛОВИЕ

«Теория вероятностей и математическая статистика» изучает свойства массовых случайных событий, способных многократно повторяться. Основное свойство случайного события есть вероятность его появления.

Теория вероятностей – математическая наука, которая опирается на первоначальную систему аксиом. Впервые законченную систему аксиом сформулировал академик А. Н. Колмогоров в своей книге «Основные понятия теории вероятностей».

По форме проявления причинно-следственных связей все законы природы и общества делятся на два класса: детерминированные и статистические. Согласно статистическим законам, будущее состояние системы определяется неоднозначно, а лишь с некоторой вероятностью.

Одной из важнейших сфер применения теории вероятностей и математическая статистика является экономика. В настоящее время трудно представить исследование и прогнозирование экономических явлений без использования эконометрического моделирования, регрессионного анализа и других методов, опирающихся на теорию вероятностей; поэтому невозможно быть грамотным экономистом без знания основ теории вероятностей и математической статистики.

При увеличении количества часов, выделяемых на самостоятельную работу студентов, большую роль играют учебные пособия. Данное учебное пособие позволяет самостоятельно подготовиться к практическим занятиям. Оно также полезно при подготовке к промежуточной аттестации.

В рассматриваемом пособии авторы стремились соблюсти чёткость и строгость изложения. Большое количество разобранных примеров и задач позволяют студенту самостоятельно или с помощью преподавателя изучить данный раздел математики. Большое внимание уделено математической статистике как наиболее востребованному способу обработки экспериментальных данных.

Учебное пособие предназначено для студентов экономических направлений образования.

Глава 1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ

1.1. Случайные события

1.1.1. Элементы комбинаторики

Комбинаторика – раздел математики, который изучает задачи выбора и расположения элементов из некоторого основного множества в соответствии с заданными правилами. Формулы и принципы комбинаторики используются в теории вероятностей для подсчёта вероятности случайных событий и, соответственно, получения законов распределения случайных величин.

Факториал. Функция $f(n)$, определённая на множестве целых неотрицательных чисел, для которой

$$f(0) = 1$$

$$f(n+1) = (n+1) \cdot f(n)$$

или

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1,$$

называется n-факториалом и обозначается $n!$.

Пример 1. Вычислить $6!$ и $9!$.

Решение. $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720;$

$$9! = 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 362880.$$

Перестановки. Каждая последовательность n различных элементов с учётом их порядка называется *перестановкой* этих элементов. Число перестановок обозначается P_n и находится по формуле

$$P_n = n!.$$

Пример 2. Сколькими способами можно расставить шесть книг на полке?

Решение. Число способов равно числу перестановок из шести элементов, то есть $P_6 = 6! = 720$.

Перестановки с повторениями. Пусть имеются N элементов, среди которых k_1 элементов одного типа, k_2 элементов другого типа, k_1 элементов

$L-$го типа $k_1+k_2+...+k_n=n.$ Число перестановок из этих N элементов перестановок с повторениями, обозначается $\overline{P_n}$ и вычисляется по формуле

$$\overline{P_n}=\frac{n!}{k_1!k_2!\cdot...k_n!}.$$

Пример 3. Десять приезжих мужчин размещаются в гостинице в двух трехместных и одном четырехместном номерах. Сколько существует способов их размещения?

*Решение***.** $N=\frac{10!}{3!\cdot 3!\cdot 4!}=4200.$

Ответ. Существует 4200 способов.

Размещение. Любой упорядоченный набор k различных элементов множества M, содержащего n элементов, называется *размещением* k элементов из n. Число размещений обозначается символом A_n^k и находится по формуле

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

Пример 4. Сколькими способами можно распределить три первых места для восьми участвующих в соревновании команд?

Решение. Нас интересует, какая из команд займёт первое, второе и третье места, то есть нам важен порядок среди отобранных трёх команд, следовательно, используем размещение. Тогда число способов найдём по формуле

$$A_8^3=\frac{8!}{(8-3)!}=\frac{8!}{5!}=\frac{8\cdot 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{5\cdot 4\cdot 3\cdot 2\cdot 1}=8\cdot 7\cdot 6=336.$$

Размещения с повторениями. Каждое размещение с повторениями из N элементов по M элементов может состоять не только из различных элементов, но из M каких угодно и как угодно повторяющихся элементов, взятых из данных N элементов.

Соединения, отличающиеся друга хотя бы порядком расположения элементов, считаются различными размещениями.

Число размещений с повторениями из n элементов по M элементов обозначается символом $\overline{A}_n^m$ и вычисляется по формуле

$$\overline{A}_n^m = n^m \qquad (2)$$

Пример 5. Для запирания сейфов и автоматических камер хранения применяют секретные замки, которых открываются лишь тогда, когда набрано некоторое тайное слово. Пусть на диск нанесено 12 букв, а секретное слово состоит из 5 букв. Сколько неудачных попыток может быть сделано человеком, не знающим секретного слова?

Решение. Общее число возможных комбинаций можно найти по формуле (2)

$$N = \overline{A}_{12}^5 = 12^5 = 248832$$

Число неудачных попыток 248832-1=248831

Ответ: 248831

Сочетание. Любое подмножество k различных элементов множества M, содержащего n элементов, называется *сочетанием*. Число сочетаний обозначается символом C_n^k и находится по формуле

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

Пример 6. Найти число способов отобрать три цветка из семи.

Решение. Так как порядок среди цветов нам не важен, то используем сочетание. Число способов найдём по формуле

$$C_7^3 = \frac{7!}{(7-3)!3!} = \frac{7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{(4\cdot 3\cdot 2\cdot 1)\cdot 3\cdot 2\cdot 1} = \frac{7\cdot 6\cdot 5}{3\cdot 2\cdot 1} = 35.$$

Сочетания с повторениями. Сочетание с повторениями из N элементов по M элементов может содержать любой элемент сколько угодно раз от 1 до M включительно или не содержать его совсем, т. е каждое сочетание из N элементов по M элементов может состоять не только из M различных элементов, но из M каких угодно и как угодно повторяющихся элементов.

Число сочетаний с повторениями из n элементов по m обозначают символом $\overline{C}_n^m$ и вычисляют по формуле

$$\overline{C}_n^m = C_{n+m-1}^m = \frac{(n+m-1)!}{m!(n-1)!}.$$

В сочетаниях с повторениями M может быть и больше N.

Пример 7. В кондитерском магазине продавались 4 сорта пирожных наполеоны, эклеры, песочник и слоеные. Сколькими способами можно купить 7 пирожных?

Решение. Число различных покупок равно числу сочетаний с повторениями из 4 по 7:

$$N = \overline{C}_4^7 = C_{4+7-1}^7 = \frac{10!}{7!3!} = 120.$$

Ответ: Из пирожных 4 сортов 7 пирожных можно выбрать 120 способами.

1.1.2. Алгебра событий

Под **событием** в теории вероятностей понимается всякий факт, который в результате опыта может произойти или не произойти.

Пример 8. Примеры событий:

– появление герба при бросании монеты;

– появление трёх гербов при трёхкратном бросании монеты;

– попадание в цель при выстреле;

– появление туза при вынимании карты из колоды;

– обнаружение объекта при одном цикле обзора радиолокационной станции;

– обрыв нити в течение часа работы ткацкого станка.

Рассматривая вышеперечисленные события, мы видим, что каждое из них обладает какой-то степенью возможности: одни – большей, другие – меньшей, причём для некоторых из этих событий мы сразу же можем решить, какое из них более, а какое менее возможно.

Введём ряд определений, необходимых для описания случайных событий.

- событие называется *случайным*, если в результате опыта оно может либо произойти, либо не произойти.
- событие называется *достоверным*, если оно обязательно происходит в результате опыта.
- событие называется *невозможным*, если оно не может произойти в данном опыте.
- события называются *несовместными*, если они не могут произойти в одном опыте.
- событие A благоприятствует событию B, если из появления события A следует, что произошло событие B.
- события образуют *полную группу*, если в результате опыта произойдёт хотя бы одно из них.
- событие C называется *суммой* событий A и B, если оно состоит в появлении события A и/или появлении события B. Сумма событий обозначается $C = A + B$.
- событие C называется *произведением* событий A и B, если оно состоит в появлении события A и появлении события B. Обозначается $C = A \cdot B$.
- событие C называется *разностью* событий A и B, если оно состоит в появлении события A и не появлении события B. Обозначается $C = A - B$.
- событие $\overline{A}$ называется *противоположным* событию A, если оно состоит в не появлении события A.
- события называются *равновозможными*, если нет объективных оснований считать, одно более возможно, чем другое.
- равновозможные, несовместные образующие полную группу события называются исходами данного опыта.
- все элементарные события, е сумме составляющие достоверное образуют пространство элементарных событий. При однократном бросании одной кости пространство элементарных событий содержит 6 элементов, при

одновременном бросании двух костей - 36 элементов (всевозможные сочетания числа очков на первой и второй кости), при попадании стрелы в мишень пространство элементарных событий содержит бесконечное множество точек мишени. На рисунках 1 и 2

- прежде, чем определить вероятность на данном пространстве элементарных событий, строят поле событий. Поле событий - это множество событий, которое включает в качестве элементов :

1. Достоверное событие,
2. Невозможное событие,
3. Все элементарные события данного пространства,
4. все события, которые на этом пространстве можно построить путем сложения (объединения) событий, путем перемножения (пересечения) событий, а также путем взятия противоположных событий от любого уже построенного.

Таким образом, никакая операция алгебры событий над заданным пространством элементарных событий не порождает события, не принадлежащего полю событий. Поле событий может содержать конечное число элементов (если конечно число элементарных событий) или бесконечное множество событий.

Наиболее строгое и общее определение понятия вероятность дал русский математик А. Н. Колмогоров. Оно гласит:

Каждому событию A из поля событий сопоставляется неотрицательное число $P(A)$, называемое вероятностью этого события и удовлетворяющее следующим аксиомам:

1. $P(A) \geq 0$;
2. $P(U) = I, U - \textit{достоверное событие}$;
3. $P(A+B) = P(A) + P(B), \textit{ если } A \textit{ и } B - \textit{несовместны.}$

Пример 9. Два шахматиста играют подряд две партии Под исходом опыта будем понимать выигрыш одного из них в l*-й* партии или ничью. Построить пространство Ω Элементарных исходов.

Решение. Обозначим события A_i – в l-й партии выиграл первый игрок, B_i — второй. C — ничью. Тогда возможные исходы игры

1. Обе партии выиграл первый игрок A_1A_2
2. Обе партии выиграл второй игрок B_1B_2
3. Обе партии закончились вничью C_1C_2.
4. В первой партии выиграл первый игрок, во второй — второй A_1B_2
5. В первой выиграл первый игрок, во второй — ничья A_1C_2
6. В первой партии победа второго игрока, во второй — первого B_1A_2
7. В первой — победа второго игрока, во второй — ничья B_1C_2
8. В первой — ничья, во второй — победа первого игрока C_1A_2
9. В первой — ничья, во второй — победа второго игрока C_1B_2

Ответ: $\Omega=\{A_1A_2, B_1B_2, C_1C_2, A_1B_2, A_1C_2, B_1A_2, B_1C_2, C_1A_2, C_1B_2\}$

Пример 10. События A, B и C означают, что взято хотя бы по одной книге из трех различных собраний сочинений, каждое из которых содержит по крайней мере три тома События A_k и B_k Означают соответственно, что из первого собрания сочинений взяты S, а из второго K томов Что означают события

а) $A+B+C$; б) ABC в) A_1+B_3 г) A_2B_2 д) $(A_1B_3+B_1A_3)C$

Решение.

1. $A+B+C$; — взята хотя бы одна книга
2. ABC — взято хотя бы по одному тому из первого, второго и третьего собрания сочинений
3. A_1+B_3 — взята одна книга из первого собрания сочинений или три книги из второго собрания сочинений, или одна из первого и три из второго собрания сочинений одновременно.
4. A_2B_2 — взято по два тома из первого и второго собрания сочинений
5. $(A_1B_3+B_1A_3)C$ — взят хотя бы один том из третьего собрания сочинений и один том из первого и три тома из второго собрания сочинений или три тома из первого и один том из второго собрания сочинений

Пример 11. Пусть $A_i(i=\overline{1,3})$ - события Ваша встреча с 1-ым другом Составьте события а) с друзьями Вы не встречались, б) Вы встречались только со вторым другом, в) с кем-то Вы не встретились, г) Вы встретились с большей частью друзей, д) у Вас состоялась встреча только с одним другом, е) Вы встретились с кем-то из первых двух друзей, а с третьим другом - нет. ж) со вторым другом Вы не встретились

Назовите события а) $\overline{A}_1 \cdot A_2 \cdot A_3 + A_1 \cdot \overline{A}_2 \cdot A_3 + \overline{A}_1 \cdot A_2 \cdot \overline{A}_3$ б) $\overline{\overline{A}_1 \cdot \overline{A}_2 \cdot \overline{A}_3}$ в) $\overline{A}_1 \cdot \overline{A}_2 + \overline{A}_1 \cdot \overline{A}_3 + \overline{A}_2 \cdot \overline{A}_3$ г) $A_1\overline{A}_2 + \overline{A}_3$ д) $A_1\overline{A}_2A_3$ е) $\overline{A_1 + A_2 + A_3}$ ж) $A_1\overline{A}_2$.

Решение. Составим события

а) Так как событие $A_i(i=\overline{1,3})$- «Ваша встреча с l-ым другом», то $\overline{A}_i(i=\overline{1,3})$ «с l-ым другом Вы не встретились». Поэтому событие «с друзьями Вы не встречались» - это совместное наступление событий $\sim A_i$. т. е. $\overline{A}_1\overline{A}_2\overline{A}_3$.

б) Слово *Только* говорит о том. что с первым и вторым другом Вы не встречались, а со вторым – да. Это $\overline{A}_1A_2\overline{A}_3$

в) Этот кто-то может быть любым из ваших друзей, поэтому событие - сумма событий $\overline{A}_i(i=\overline{1,3})$ т.е. $\overline{A}_1 + \overline{A}_2 + \overline{A}_3$.

г) Так как друзей трое, а большая часть - это более половины, то Вы встретились, по крайней мере, с двумя друзьями, поэтому событие - сумма событий $A_j \cdot A_i(j,\ i=\overline{1,3})$ и $i=j$.. т. е $A_1A_2 + A_1A_3 + A_2A_3$.

д) Этим одним другом может быть любой из Ваших трех друзей, поэтому это событие есть сумма таких событий «Вы встретились только с первым другом» или «встретились только со вторым», или «встретились только с третьим», т. е. $A_1 \cdot \overline{A}_2 \cdot \overline{A}_3 + \overline{A}_1 \cdot A_2 \cdot \overline{A}_3 + \overline{A}_1 \cdot A_2 \cdot \overline{A}_3$.

е) Встреча с кем-то из первых двух друзей - это встреча либо с первым другом, либо со вторым (а может быть и с обоими), т е это сумма $A_1 + A_2$ и в то же время не встретились с третьим. Поэтому ответ $(A_1 + A_2)\overline{A}_3$

ж) Так как A_2- «встреча со вторым другом», то $\overline{A}_2$ «встречи со вторым другом не было». Так как про других друзей ничего не говорится, то не надо думать про встречи с ними.

Назовем события

а) Вы не встретились только с одним другом (или Вы встретились только с двумя).

б) Событие $\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$ -«ни с кем Вы не встретились», а событие $\overline{\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}}$ противоположно событию $\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$ (отрицание этого события). Поэтому ответ: встречи были $(A_1 + A_2 + A_3)$ (с кем-то Вы встретились). Итак. $\overline{\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}} = A_1 + A_2 + A_3.$

в) С двумя друзьями Вы не встречались (с большей частью своих друзей Вы не встречались).

г) С первым другом Вы встретились, а с кем-то из остальных - нет.

д) Вы не встретились только со вторым другом (или у Вас была встреча только с первым и третьим другом)

е) Так как $A_1 + A_2 + A_3$ событие «Вы с кем-то встречались», то событие $\overline{A_1 + A_2 + A_3}$ - противоположное (отрицание этого события - «Вы ни с кем не встречались», т. е. $\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$. Итак, $\overline{A_1 + A_2 + A_3} = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$.

1.1.3. Классическое определение вероятности

Для определения вероятности события используем следующие понятия.

Вероятностью события называется численная мера степени объективной возможности появления этого события.

Вероятностью (классическое определение вероятности) события A называется отношение числа благоприятствующих исходов к общему числу исходов данного опыта.

Вероятность события A обозначается $p(A)$.

Тогда

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

где m – число благоприятных для появления события A исходов; n – число всевозможных исходов опыта.

Основные свойства вероятности:

1. Вероятность достоверного события равна единице.

2. Вероятность невозможного события равна нулю.

3. Для любого события A его вероятность заключена в интервале $0 \le p(A) \le 1$.

4. Вероятность наступления противоположного события $\overline{A}$ равна разности между единицей и вероятностью события A, то есть $p(\overline{A}) = 1 - p(A)$.

Пример 12. Из урны, содержащей 12 чёрных и 8 белых шаров, наудачу вынуто два шара. Найти вероятность того, что они разного цвета.

Решение. Обозначим событие A – шары разного цвета, тогда по определению искомая вероятность равна

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

где m – число способов отобрать один чёрный шар из 12 и один белый шар из 8; n – число всевозможных способов отобрать два шара любого цвета из двадцати.

Найдём число способов отобрать два шара любого цвета из двадцати:

$$n = C_{20}^{2} = \frac{20!}{18! \cdot 2!} = 190.$$

Число благоприятных исходов равно числу способов отобрать один белый шар из 8 и один чёрный шар из 12. Общее число благоприятных исходов равно

$$m = C_{8}^{1} \cdot C_{12}^{1} = \frac{8!}{7! \cdot 1!} \cdot \frac{12!}{11! \cdot 1!} = 96.$$

Поэтому искомая вероятность равна

$$p = \frac{m}{n} = \frac{96}{190} = \frac{48}{95}.$$

Пример 13. Из колоды карт наудачу вынуто две. Найти вероятность того, что они обе бубновой масти. Колода содержит 36 карт.

Решение. Обозначим событие A – из колоды карт наудачу вынуть две бубновые карты.

По классическому определению вероятности имеем:

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

где m – число способов, которыми можно вынуть из колоды две карты бубновой масти; n – общее число всевозможных исходов. То есть

$$m=C_9^2=\frac{9!}{2!\cdot 7!}=36; \qquad n=C_{36}^2=\frac{36!}{2!\cdot 34!}=630.$$

Тогда $p(A)=\frac{36}{630}=\frac{2}{35}$.

Пример 14. 1 сентября на первом курсе одного из факультетов запланированы по расписанию три лекции из 10 различных предметов Студент, не успевший ознакомиться с расписанием, пытается его угадать. Какова вероятность успеха в данном эксперименте, если считать, что любое расписание из трех предметов равновозможной.

Решение. Студенту необходимо из 10 лекций, которые могут быть поставлены в расписание, причем в определенном порядке, выбрать три. Следовательно, число всех возможных исходов испытания равно числу размещений из 10 по 3. т е

$$n=A_{10}^3=\frac{10!}{7!}=10\cdot 9\cdot 8=720.$$

Благоприятный же случай только один, т е. $M=1.$ Искомая вероятность будет равна

$$P=\frac{m}{n}=\frac{1}{720}\approx 0,0014.$$

Ответ: $P=\frac{1}{720}\approx 0,0014.$

Пример 15. В подъезде дома установили замок с кодом. Дверь автоматически отпирается, если в определенной последовательности набрать три цифры из возможных десяти. Некто вошел в подъезд и, не зная, кода стал наудачу пробовать различные комбинации из трех цифр. На каждую попытку он тратит 15 секунд. Какова вероятность события А= {вошедшему удастся открыть дверь за один час)?

Решение. Так как цифры, входящие в набираемый номер, могут повторяться и порядок их набора играет существенную роль, то мы приходим к схеме размещений с повторениями. Число возможных вариантов набора трех цифр из 10 возможных равно $\overline{A}_{10}^{3} = 10^3$. За один час, тратя на набор комбинации 15 секунд, можно набрать 240 различных комбинаций, т. е $M = 240$. Искомая вероятность

$$P = \frac{m}{n} = \frac{240}{10^3} = 0,24$$

Ответ: $P = 0,24.$

Пример 16. На полке стоят 15 книг. 5 из них в переплете Берут наудачу три книги Какова вероятность того, что все три книги в переплете?

Решение. Опыт состоит в том, что из 15 книг отбирают 3. причем в каком порядке они отобраны, роли не играет. Следовательно, число возможных способов выбора будет равно числу сочетаний из 15 по 3. т е.

$$n = C_{15}^{3} = \frac{15!}{3!\cdot 12!} = \frac{13\cdot 14\cdot 15}{6} = 455.$$

Число благоприятных случаев будет равно числу сочетаний из 5 по 3, т. е.

$m = C_5^3 = \dfrac{5!}{3!\cdot 2!} = 10.$ Искомая вероятность $P = \dfrac{m}{n} = \dfrac{10}{455} \approx 0,022.$

Ответ: $P \approx 0,022.$

Пример 17. Десять приезжих мужчин, среди которых Петров и Иванов, размещаются в гостинице в двух трехместных и одном четырехместном номерах. Какова вероятность события A, состоящего в том, что Петров и Иванов попадут в четырехместный номер?

Решение. Число всех возможных размещений 10 человек в двух трехместных и одном четырехместном номере равно числу перестановок из десяти элементов, среди которых 3 одного вида. 3 другого и 4 третьего, т. е.

$$n = \overline{P}_{10}(3;3;4) == \frac{10!}{3!\cdot 3!\cdot 4!} = 4200.$$

После того как Иванов и Петров будут размещены в четырехместном номере, остальные 8 человек должны быть размещены в двух трехместных и на оставшиеся два свободных места в четырехместном номере, это можно будет сделать следующим образом:

$$m = \overline{P}_8(3;3;2) == \frac{8!}{3!\cdot 3!\cdot 2!} = 560.$$

Искомая вероятность $P = \frac{m}{n} = \frac{560}{4200} \approx 0{,}133.$

Ответ: $P \approx 0{,}133.$

1.1.4. Геометрическая вероятность

Геометрической вероятностью события A называется отношение меры области благоприятных исходов к мере области всевозможных исходов:

$$p(A) = \frac{S_{\text{бл}}}{S},$$

где $S_{\text{бл}}$ – площадь области благоприятных исходов; S – площадь области всевозможных исходов.

Пример 18. (Задача о встрече Рис.1.) Два студента условились встретиться в определённом месте между 12-00 и 12-30 часами. Пришедший первым ждёт второго в течение 10 минут, после чего уходит. Найти вероятность того, что встреча состоится, если каждый студент наудачу выбирает момент своего прихода (в промежутке от 12-00 до 12-30 часов).

Решение. Обозначим событие A – студенты встретятся, тогда противоположное событие $\overline{A}$ – студенты не встретятся. Обозначим момент прихода одного из них x (минуты), момент прихода другого y (минуты). Для того чтобы встреча произошла, необходимо и достаточно, чтобы выполнялось условие

$$|x - y| \le 10.$$

Будем изображать x и y как декартовые координаты точек плоскости. Координаты $(x,\ y)$ однозначно определяют время прихода обоих студентов.

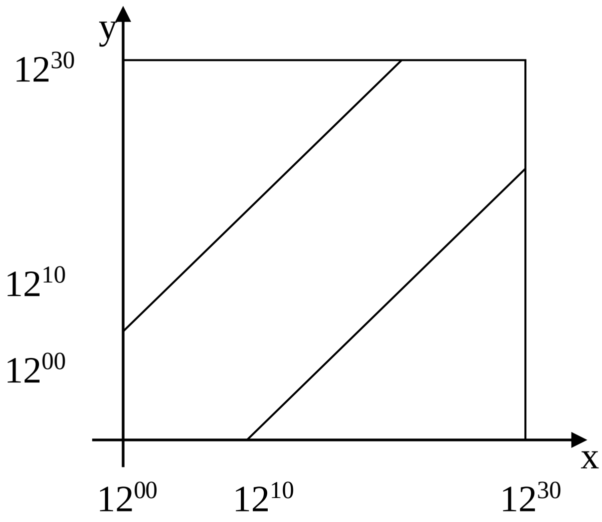

Рис. 1. Задача о встрече.

Все возможные исходы изображены точками фигуры, ограниченной квадратом, со стороной 30, площадь квадрата равна $S = 30^2 = 900$.

Встреча состоится, если выполнено неравенство $|x - y| \le 10$. Данное неравенство равносильно системе

$$\begin{cases} x - y \le 10; \\ x - y \ge -10; \end{cases}$$

построим две эти прямые. Событие A – студенты встретятся удовлетворяет системе неравенств

$$\begin{cases} x - y \le 10; \\ x - y \ge -10; \\ 0 \le x \le 30; \\ 0 \le y \le 30. \end{cases}$$

Площадь данной области равна

$$S_A = 30^2 - 20^2 = 900 - 400 = 500.$$

Тогда искомая вероятность равна

$$p(A) = \frac{500}{900} = \frac{5}{9}.$$

1.1.5. Теоремы сложения

Теорема 1. Для несовместных событий вероятность появления суммы событий равна сумме вероятностей. То есть

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

Пример 19. Вероятности получить на экзамене 5, 4, 3 соответственно равны 0,2; 0,3; 0,3. Найти вероятность сдачи экзамена.

Решение. Обозначим события:

A – студент сдал экзамен;

B – студент сдал экзамен на 5;

C – студент сдал экзамен на 4;

K – студент сдал экзамен на 3.

По условию задачи имеем: $p(B)=0{,}2$; $p(C)=0{,}3$; $p(K)=0{,}3$. Применяем теорему сложения в силу несовместности событий B, C, K, имеем:

$$p(A)=p(B+C+K)=p(B)+p(C)+p(K)=0{,}2+0{,}3+0{,}3.$$

Теорема доказана.

Теорема 2. Для любых событий вероятность появления суммы этих событий равна сумме вероятностей этих событий без вероятности их совместного появления. То есть

$$p(A+B)=p(A)+p(B)-p(A\cdot B).$$

1.1.6. Теоремы умножения

Два события называются *независимыми*, если вероятность появления одного из них не зависит от того, произошло или не произошло второе событие.

Несколько событий называются *независимыми в совокупности*, если любая комбинация из них независима.

События называются *зависимыми*, если появление или непоявление одного из них изменяет вероятность появления другого.

Вероятность события B, вычисленная в предположении осуществления события A, называется *условной вероятностью* события B и обозначается $p_A(B)$ или $p(B/A)$.

Пример 20. Из урны, содержащей 2 белых и 3 чёрных шара, наудачу последовательно извлекают два шара. Найти вероятность того, что второй извлечённый шар белый, если известно, что первый извлечённый шар чёрный.

Решение. Обозначим события:

A – первый извлечённый шар чёрный;

B – второй извлечённый шар белый.

Для нахождения искомой вероятности используем классическое определение:

$$p_A(B) = \frac{m}{n},$$

где m – число благоприятных исходов, равное числу оставшихся после первого извлечения белых шаров, то есть $m = 2$; n – число всевозможных оставшихся шаров, то есть $n = 4$. Тогда искомая вероятность равна $p_A(B) = \frac{2}{4} = \frac{1}{2}$.

Теорема 3. Вероятность произведения двух событий равна произведению вероятности одного из них на условную вероятность другого, вычисленную при условии, что первое событие произошло. То есть

$$p(A \cdot B) = p(A) \cdot p_A(B) = p(B) \cdot p_B(A).$$

Следствие 1. Если появление события A не зависит от события B, то появление события B не зависит от события A.

Следствие 2. Вероятность произведения двух независимых событий равна произведению вероятности этих событий.

Пример 21. Известно, что 85 % готовой продукции цеха является стандартной. Вероятность того, что стандартная деталь отличного качества, равна 0,51. Найти вероятность того, что наудачу взятое изделие окажется отличного качества.

Решение. Пусть A – событие, означающее, что взятое наудачу изделие стандартное, B – событие, означающее, что изделие отличного качества. Изделие может быть отличного качества, если оно стандартное. Поэтому из условия задачи следует $p(A)=0{,}85$, а $p_A(B)=0{,}51$. Тогда искомая вероятность равна $p(A\cdot B)=p(A)\cdot p_A(B)=0{,}85\cdot 0{,}51=0{,}4335.$

Пример 22. Решить задачу, применяя теоремы сложения и умножения Мастер обслуживает 3 станка, работающих независимо друг от друга. Вероятность того, что первый станок потребует внимания рабочего в течение смены, равна 0,4, второй -0.6, третий -0,3. Найти вероятность того, что в течение смены: а) ни один станок не потребует внимания мастера, б) ровно 1 станок потребует внимания мастера.

Решение. Вводим базовые независимые события A_1= (Станок i потребовал внимания рабочего в течение смены), $i=1,2,3$. По условию выписываем вероятности: $p_1=0,4,\ p_2=0,6,\ p_3=0,3.$ Тогда $q_1=0,6,\ q_2=0,4,\ q_3=0,7.$ Найдем вероятность события $X=$ (Ни один станок не потребует внимания в течение смены):

$$P(X)=P(\overline{A_1}\,\overline{A_2}\,\overline{A_3})=q_1q_2q_3=0,6\cdot 0,4\cdot 0,7=0{,}168.$$

Найдем вероятность события Z- (Ровно один станок потребует внимания в течение смены):

$$P(Z)=P(A_1)P(\overline{A_2})P(\overline{A_3})+P(\overline{A_1})P(A_2)P(\overline{A_3})+P(\overline{A_1})P(\overline{A_2})P(A_3)=$$

$$=p_1q_2q_3+q_1p_2q_3+q_1q_2p_3=0,436$$

Пример 23. Студент разыскивает нужную ему формулу в трех справочниках. Вероятности того, что формула содержится в первом, втором и третьем справочниках равны 0,6; 0,7 и 0,8. Найти вероятности того, что формула содержится 1) только в одном справочнике; 2) только в двух справочниках; 3) во всех трех справочниках.

Решение.

A - формула содержится в первом справочнике;

В - формула содержится во втором справочнике;

С - формула содержится в третьем справочнике. i

Воспользуемся теоремами сложения и умножения вероятностей.

1. $P(A\overline{B}\overline{C}+\overline{A}C\overline{B}+\overline{A}\overline{C}B)$
2. $P(AB\overline{C}+A\overline{B}C+\overline{A}BC)=0,6\cdot 0,7\cdot 0,2+0,6\cdot 0,3\cdot 0,8+0,4\cdot 0,7\cdot 0,8=0,452$
3. $P(ABC)=0,6\cdot 0,7\cdot 0,8=0,336.$

Вероятность наступления хотя бы одного события

Пусть в результате испытания могут появиться n событий, независимых в совокупности, либо некоторые из них (в частности, только одно или ни одного), причем вероятности появления каждого из событий известны. Как найти вероятность того, что наступит хотя бы одно из этих событий?

Например, если в результате испытания могут появиться три события, то появление хотя бы одного из этих событий означает наступление либо одного, либо двух, либо трех событий. Ответ на поставленный вопрос дает следующая теорема.

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

$$P(A)=1-P(\overline{A_1})P(\overline{A_2})...P(\overline{A_n})=1-q_1q_2...q_n$$

Если события $A_1, A_2, ..., A_n$ имеют одинаковую вероятность p, то формула принимает простой вид:

$$P(A)=1-(1-p)^n=1-q^n.$$

Общая методика для решения задач, в которых встречается фраза "хотя бы один" такая:

1. Выписать исходное событие $\boldsymbol{A}$ = (Вероятность того, что ... хотя бы ...).
2. Сформулировать противоположное событие $\overline{A}$.
3. Найти вероятность события $P(\overline{A})$.
4. Найти искомую вероятность по формуле $P(A)=1-P(\overline{A})$.

А теперь разберем ее на примерах.

Пример 24. В урне 2 белых, 3 черных и 5 красных шаров. Три шара вынимают наугад. Найти вероятность того, что среди вынутых шаров хотя бы два будут разного цвета.

Решение.

1. Записываем событие $A=$ (Среди вынутых 3 шаров **хотя бы два** разного цвета). То есть, например, "2 красных шара и 1 белый", или "1 белый, 1 черный, 1 красный", или "2 черных, 1 красный" и так далее, вариантов многовато. Попробуем правило перехода к противоположному событию.

2. Тогда противоположное событие формулируется так $\overline{A}=$ (Все три шара одного цвета) - (Выбраны 3 черных шара или 3 красных шара) - всего 2 варианта получилось, значит, этот способ решения упрощает вычисления. Кстати, все шары белого цвета не могут быть выбраны, так как их всего 2, а вынимается 3 шара.

3. Общее число исходов (способов выбрать любые 3 шара из $2+3+5=10$ шаров) равно $n=C_{10}^{3}=120$.

Найдем числоблагоприятствующих событию исходов $m=C_3^3+C_5^3=1+10=11$ число способов выбрать или 3 черных шара (из 3), или 3 красных шара (из 5).Тогда

$$P(\overline{A})=\frac{m}{n}=\frac{11}{120}$$

4. Искомая вероятность: $P(A)=1-P(\overline{A})=1-\frac{11}{120}=0,908.$

Ответ: P = 0,908.

1.1.7. Формула полной вероятности

Теорема 4. Если событие A может наступить только при условии появления одного из событий $H_1, H_2, ..., H_n$, образующих полную группу несовместных событий, то вероятность события A равна сумме произведений вероятностей каждого из событий $H_1, H_2, ..., H_n$ на соответствующую условную вероятность события A:

$$p(A)=\sum_{k=1}^{n} p(H_k)\cdot p_{H_k}(A).$$

Эта формула называется формулой полной вероятности, события $H_1, H_2, ..., H_n$ – гипотезы, причём сумма вероятностей гипотез равна единице, то есть

$$\sum_{k=1}^{n} p(H_k)=1.$$

Пример 25. Три завода производят одинаковые изделия, которые поступают на один склад, причём первый завод производит 30 %, второй – 20 % и третий – 50 % от всех поступивших на склад деталей. Для первого завода брак составляет 5 %, для второго – 8 % и для третьего – 10 %. Найти вероятность того, что наудачу взятая со склада деталь окажется бракованной.

Решение. Обозначим события:

A – деталь бракованная;

H_1 – деталь произведена на первом заводе;

H_2 – деталь произведена на втором заводе;

H_3 – деталь произведена на третьем заводе.

По условию задачи можем записать:

$$p(H_1)=0{,}3;$$

$$p(H_2)=0{,}2;$$

$$p(H_3)=0{,}5;$$

$$p_{H_1}(A)=0{,}05;$$

$$p_{H_2}(A)=0{,}08;$$

$$p_{H_3}(A)=0{,}1.$$

Тогда искомую вероятность необходимо вычислять по формуле полной вероятности:

$$p(A)=\sum_{k=1}^{3} p(H_k)\cdot p_{H_k}(A)=0{,}3\cdot 0{,}05+0{,}2\cdot 0{,}08+0{,}5\cdot 0{,}1=0{,}081.$$

Пример 26. B группе спортсменов лыжников в $2 см$ больше, чем бегунов, а бегунов в 3 раза больше, чем велосипедистов. Вероятность выполнить норму для лыжника 0.9, для бегуна 0.75, для велосипедиста - 0.8.
Найти вероятность того, что спортсмен, выбранный наугад, выполнит норму.

Решение. Введем полную группу гипотез

$H_1 =$ (Спортсмен - лыжник),

$H_2 =$ (Спортсмен - велосипедист),

$H_3 =$ (Спортсмен - бегун).

Найдем вероятности гипотез по классическому определению вероятности. Пусть велосипедистов x, тогда бегунов будет $3x$, а лыжников $6x$. Получаем

$$P(H_1) = \frac{6x}{6x+3x+x} = 0,6$$

$$P(H_2) = \frac{x}{6x+3x+x} = 0,1$$

$$P(H_1) = \frac{3x}{6x+3x+x} = 0,3.$$

Введем событие $A =$ (Спортсмен выполнит норму). Известны вероятности

$$P(A/H_1) = 0,9,\ P(A/H_2) = 0,8,\ P(A/H_3) = 0,75.$$

Тогда вероятность события A найдем по формуле полной вероятности

$$P(A) = P(A/H_1)P(H_1) + P(A/H_2)P(H_2) + P(A/H_3)P(H_3) = 0,845.$$

1.1.8. Формула Байеса

Формула Байеса применяется при решении практических задач в том случае, когда событие A, появляющееся совместно с каким-либо из событий $H_1, H_2, ..., H_n$, произошло и требуется произвести количественную переоценку вероятностей событий $H_1, H_2, ..., H_n$ (события $H_1, H_2, ..., H_n$ образуют полную группу несовместных событий).

Априорные вероятности: $p(H_1),\ p(H_2),\ p(H_n)$ известны. Требуется вычислить апостериорные вероятности: $p_A(H_1),\ p_A(H_2),\dots, p_A(H_n)$.

Пусть событие A может наступить при условии появления одной из гипотез $H_1, H_2, ..., H_n$. Вероятность совместного появления события A с одной из гипотез H_m по теореме умножения равна

$$p(A \cdot H_m) = p(A) \cdot p_A(H_m) = p(H_m) \cdot p_{H_m}(A),$$

отсюда получаем

$$p_A(H_m) = \frac{p(H_m) \cdot p_{H_m}(A)}{p(A)}$$

или

$$p_A(H_m) = \frac{p(H_m) \cdot p_{H_m}(A)}{p(H_1) \cdot p_{H_1}(A) + p(H_2) \cdot p_{H_2}(A) + ... + p(H_n) \cdot p_{H_n}(A)}.$$

Полученные формулы носят название *формул Байеса.*

Пример 27. В первой группе 20 студентов, во второй группе 25 студентов, а в третьей группе 15 студентов. Вероятность сдать экзамен на отлично для студентов первой группы равна 0,6, для студентов второй – 0,3, для третьей – 0,4. Наудачу взятый студент сдал экзамен на отлично. Найти вероятность того, что он из третьей группы.

Решение. Обозначим события:

A – наудачу взятый студент сдал экзамен на отлично;

H_1 – студент из первой группы;

H_2 – студент из второй группы;

H_3 – студент из третьей группы.

Определим вероятности:

$$p(H_1) = \frac{20}{20 + 25 + 15} = \frac{20}{60} = \frac{1}{3};$$

$$p(H_2) = \frac{25}{20 + 25 + 15} = \frac{25}{60} = \frac{5}{12};$$

$$p(H_3) = \frac{15}{20 + 25 + 15} = \frac{15}{60} = \frac{1}{4};$$

$$p_{H_1}(A) = 0{,}6;$$

$$p_{H_2}(A) = 0{,}3;$$

$$p_{H_3}(A) = 0{,}4.$$

Тогда искомая вероятность $p_A(H_3)$ по формуле Байеса равна

$$p_A(H_3) = \frac{p(H_3)\cdot p_{H_3}(A)}{p(H_1)\cdot p_{H_1}(A) + p(H_2)\cdot p_{H_2}(A) + p(H_3)\cdot p_{H_3}(A)} =$$

$$= \frac{\frac{1}{4}\cdot 0{,}4}{\frac{1}{3}\cdot 0{,}6 + \frac{5}{12}\cdot 0{,}3 + \frac{1}{4}\cdot 0{,}4} = \frac{4}{17}.$$

Пример 28. На трех станках автоматах обрабатываются однотипные детали, поступающие после обработки на общий конвейер. Первый станок дает 2% брака, второй - 7%. третий - 10%. Производительность первого станка в 3 раза больше производительности второго, а третьего - в 2 раза меньше, чем второго.

а) Каков процент брака на конвейере?

б) Каковы доли деталей каждого станка среди бракованных деталей на конвейере?

Решение. Возьмем с конвейера наудачу одну деталь и рассмотрим событие А - деталь бракованная. Оно связано с гипотезами относительно того, где была обработана эта деталь: H_k - взятая наудачу деталь обработана на $k-$ом станке, $k = 1, 2, 3$.

Условные вероятности (в условии задачи они даны в форме процентов):

$$P(A/H_1) = 0{,}02,\ P(A/H_2) = 0{,}07,\ P(A/H_3) = 0{,}1.$$

Зависимости между производительностями станков означают следующее:

$$P(H_1) = 3P(H_2),\ P(H_3) = 0{,}5P(H_2).$$

Атак как гипотезы образуют полную группу, то $P(H_1) + P(H_2) + P(H_3) = 1$.

Решив полученную систему уравнений, найдем:

$$P(H_1) = 6/9,\ P(H_2) = 2/9,\ P(H_3) = 1/9.$$

а) Полная вероятность того, что взятая наудачу с конвейера деталь - бракованная:

$$P(A) = \sum P(A / H_i) P(H_i) = 0,04.$$

Другими словами, в массе деталей, сходящих с конвейера, брак составляет 4%.

б) Пусть известно, что взятая наудачу деталь - бракованная. Пользуясь формулой Байеса, найдем условные вероятности гипотез:

$$P\left(H_1 / A\right) = \frac{P(A / H_1) P(H_1)}{P(A)} = \frac{6/9 \cdot 0,02}{0,04} = 0,33$$

$$P\left(H_2 / A\right) = \frac{P(A / H_2) P(H_2)}{P(A)} = \frac{2/9 \cdot 0,07}{0,04} = 0,39$$

$$P\left(H_3 / A\right) = \frac{P(A / H_3) P(H_3)}{P(A)} = \frac{1/9 \cdot 0,1}{0,04} = 0,28.$$

Таким образом, в общей массе бракованных деталей на конвейере доля первого станка составляет 33%. второго - 39%. третьего - 28%.

Рассмотрим две гипотезы

B_1 – наудачу взятое изделие будет из 1-й партии.

B_2 –наудачу взятое изделие будет из 2-й партии

Всего $4000 + 6000 = 10000$ изделий на складе. По классическому определению

$P(B_1) = \dfrac{4000}{10000} = 0,4,$ $P(B_2) = \dfrac{6000}{10000} = 0,6.$ Контроль $P(B_1) + P(B_2) = 0,4 + 0,6 = 1.$

Рассмотрим зависимое событие $A =$ наудачу взятое со склада изделие будет стандартным.

В первой партии 100% - 20% = 80% стандартных изделий, поэтому $P_{B_1}(A) = \dfrac{80}{100} = 0,8 -$ вероятность того, что наудачу взятое на складе изделие будет стандартным при условии, что оно принадлежит 1-й партии.

Аналогично, во второй партии 100% -10% = 90% стандартных изделий и $P_{B_2}(A) = \dfrac{90}{100} = 0,9 -$ вероятность того, что наудачу взятое на складе изделие будет стандартный при условии, что оно принадлежит 2-й партии. По формуле полной вероятности $P(A) = P(B_1) P_{B_1}(A) + P(B_2) P_{B_2}(A) = 0,4 \cdot 0,8 + 0,6 \cdot 0,9 = 0,86$ - вероятность того, что на удачу взятое на складе изделие будет стандартным.

Часть вторая. Пусть наудачу взятое со склада изделие оказалось стандартным. Эта фраза прямо прописана в условии, и она констатирует тот факт, что событие A произошло. По формулам Байеса

а) $P_A(B_1)=\dfrac{P(B_1)P_{B_1}(A)}{P(A)}=\dfrac{0,4\cdot 0,8}{0,86}\approx 0,37$ - вероятность того, что выбранное стандартное изделие принадлежит 1-й партии.

б) $P_A(B_2)=\dfrac{P(B_2)P_{B_2}(A)}{P(A)}=\dfrac{0,6\cdot 0,9}{0,86}\approx 0,63$-вероятность того, что выбранное стандартное изделие принадлежит 2-й партии.

После *переоценки* гипотезы, B_1, B_2.разумеется, по-прежнему образуют полную группу $P_A(B_1)+P_A(B_2)=\dfrac{16}{43}+\dfrac{27}{43}=\dfrac{43}{43}=1.$

Ответ: $a)\ \dfrac{16}{43}\approx 0,37,\ б)\dfrac{27}{43}\approx 0,63.$

1.1.9. Схема независимых испытаний. Формула Бернулли

На практике часто приходится сталкиваться с задачами, которые можно представить в виде многократно повторяющихся испытаний, в результате каждого из которых может появиться или не появиться событие A. При этом представляет интерес не исход каждого отдельного испытания, а общее число появлений события A в результате определённого количества испытаний. В подобных задачах нужно уметь определять вероятность любого числа m появлений события A в результате общего числа n испытаний. Рассмотрим случай независимых испытаний. Испытания называются *независимыми*, если результат одного испытания не зависит от результатов других испытаний и вероятность появления события A в каждом испытании постоянна.

Пусть происходит n независимых испытаний, в каждом из которых событие A может либо появиться с вероятностью p, либо не появиться с вероятностью $q=1-p$.

Рассмотрим некоторое событие B_m, состоящее в том, что событие A в этих n испытаниях появилось ровно m раз и, следовательно, не появилось $(n-m)$ раз. Обозначим через A_k $(k=1, 2, 3, ..., n)$ появление события A в k-м испытании, через $\overline{A_k}$ – не появление события A в k-м испытании. Тогда по условию

$$p(A_1)=p(A_2)=...=p(A_n)=p$$

$$p(\overline{A_1})=p(\overline{A_2})=...=p(\overline{A_n})=q.$$

Событие может появиться m раз в различных последовательностях или комбинациях, чередуясь с противоположным событием $\overline{A}$. Число возможных комбинаций такого рода равно числу сочетаний из n элементов по m, то есть C_n^m. Следовательно, событие B_m можно представить в виде суммы различных комбинаций, несовместных между собой, причём число слагаемых будет равно C_n^m.

В каждое слагаемое событие A входит m раз, а событие $\overline{A}$ входит $(n-m)$ раз. Вероятность каждой такой последовательности по теореме умножения равна $(p^m \cdot q^{n-m})$. Так как общее число таких последовательностей равно C_n^m, то, используя теорему сложения для несовместных событий, получим вероятность события B_m, равную $C_n^m \cdot p^m \cdot q^{n-m}$ или

$$p_n(m)=C_n^m \cdot p^m \cdot q^{n-m}.$$

Данная формула носит название *формулы Бернулли*.

Пример 29. Найти вероятность того, что при пяти бросаниях монеты герб появится ровно один раз.

Решение. Вероятность появления герба при одном бросании монеты равна $p=\frac{1}{2}$, вероятность его не появления соответственно равна $q=1-p=\frac{1}{2}$.

Тогда вероятность появления одного герба при пяти бросаниях по формуле Бернулли равна $p_5(1) = C_5^1 \cdot \left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^{5-1} = \frac{5!}{4! \cdot 1!} \cdot \frac{1}{2} \cdot \frac{1}{16} = \frac{5}{32}$.

Пример 30. Вася выигрывает у Пети в «Quake» в среднем три партии из четырех. Найти вероятность того, что он выиграет равно три из четырех сыгранных партий.

Решение: Вероятность успеха равна0,75. По формуле Бернулли искомая вероятность равна

$$P(S_4 = 3) = C_4^3 (0.75)^3 \cdot 0.25 = \frac{27}{64}$$

Пример 31. Прибор составлен из 10 блоков, надежность каждого из них 0,8. Блоки могут выходить из строя независимо друг от друга. Найти вероятность того, что:

а) откажут два блока;

б) откажет хотя бы один блок;

в) откажут не менее двух блоков.

Найти наиболее вероятное количество блоков, которые выйдут из строя.

Решения: Пусть событие А – отказ блока. Тогда

$$P(A) = p = 1 - 0.8 = 0.2$$

Поэтому $q = 0.8$. По условию задачи $n = 10$. Применив формулу Бернулли и следствие из нее, получим:

а) $P_{10}(2) = C_{10}^2 p^2 q^2 = 45 \cdot (0.2)^2 \cdot (0.8)^2 \approx 0.302$;

б) $P_{10}(1 \le m \le 10) = 1 - q^{10} = 1 - (0.8)^{10} \approx 0.893$;

в) $P_{10}(2 \le m \le 10) = 1 - [P_{10}(0) + P_{10}(1)] = 1 - C_{10}^0 \cdot (0.2)^1 \cdot (0.8)^9 \approx 0.624$.

1.1.10. Наивероятнейшее число появления событий

Наивероятнейшим числом появления события A в n независимых испытаниях называется число m_0, для которого вероятность, соответствующая этому числу, не меньше вероятности каждого из остальных возможных чисел появления события A. Наивероятнейшее число m_0 удовлетворяет неравенству

$$n \cdot p - q \le m_0 \le n \cdot p + p.$$

Рассчитаем длину интервала, определяемого данным неравенством:

$$(n \cdot p + p) - (n \cdot p - q) = p + q = 1.$$

Так как длина интервала равна единице, а событие может произойти в n испытаниях только целое число раз, то следует иметь в виду, что:

1) если $(n \cdot p - q)$ – целое число, то существуют два значения наивероятнейшего числа:

$$m_0^1 = n \cdot p - q \text{ и } m_0^2 = n \cdot p + p;$$

2) если $(n \cdot p - q)$ – дробное число, то существует единственное целое число, удовлетворяющее полученному неравенству.

Пример 32. Отдел технического контроля проверяет партию из 20 деталей. Вероятность того, что деталь стандартная, равна 0,88. Найти наивероятнейшее число деталей, которые будут признаны стандартными.

Решение. По условию задачи имеем:

$$n = 20;$$
$$p = 0{,}88;$$
$$q = 1 - p = 1 - 0{,}88 = 0{,}12.$$

Подставляя в неравенство, получим:

$$20 \cdot 0{,}88 - 0{,}12 \le m_0 \le 20 \cdot 0{,}88 + 0{,}88$$

или

$$17{,}48 \le m_0 \le 18{,}48.$$

Единственное целое число, принадлежащее этому интервалу, равно 18. Следовательно, наивероятнейшее число стандартных деталей в партии 18.

1.1.11. Асимптотическая формула Лапласа

При большом числе испытаний пользоваться формулой Бернулли достаточно трудно. Поэтому при большом числе испытаний $(n \ge 20)$ пользуются целым рядом приблизительных формул.

Производится n независимых испытаний, в каждом из которых может появиться или не появиться событие A. Вероятность наступления события A в каждом испытании постоянна и равна p $(0 < p < 1)$. Если число испытаний велико, то вероятность наступления события A ровно m раз (безразлично, в какой последовательности) приближённо может быть найдена по асимптотической формуле Лапласа:

$$p_n(m) \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{m - n \cdot p}{\sqrt{npq}}\right),$$

где $q = 1 - p$; $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Свойства функции $\varphi(x)$:

1. Функция $\varphi(x)$ чётная, то есть $\varphi(-x) = \varphi(x)$.
2. Для $x > 4$ значение функции $\varphi(x)$ близко к нулю.
3. Значения функции $\varphi(x)$ приведены в приложении, табл. 1.

Пример 33. Вероятность того, что станок-автомат произведёт стандартную деталь, равна $p = 0{,}9$. Найти вероятность того, что из 900 изготовленных деталей стандартными будут 804 детали.

Решение. По условию задачи имеем:

$$n = 900;\ m = 804;\ q = 1 - p = 0{,}1,$$

тогда

$$x = \frac{m - n \cdot p}{\sqrt{npq}} = \frac{804 - 900 \cdot 0{,}9}{\sqrt{900 \cdot 0{,}9 \cdot 0{,}1}} = -\frac{6}{9};$$

$$\varphi\left(-\frac{6}{9}\right) \approx 0{,}3187;$$

$$p_{900}(804) \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{m - n \cdot p}{\sqrt{npq}}\right) \approx \frac{1}{9} \cdot 0{,}3187 \approx 0{,}0354.$$

Пример 34. Игральный кубик подбрасывают 800 раз. Какая вероятность того, что количество очков, кратное трём, появится 267 раз?

Решение. В этой задаче количество испытаний $n=800$ довольно большое. Вероятность p того, что при подбрасывании кубика выпадает число, кратное трем, равняется $\frac{1}{3}$ - и постоянная для всех испытаний, $q=1-p=\frac{2}{3}$**.** Поэтому для вычисления вероятности воспользуемся локальной теоремой Муавра—Лапласа. Для этого найдём *x:*

$$x=\frac{267-800\cdot\frac{1}{3}}{\frac{40}{3}}=0.025$$

Итак,

$$P_{800}(267)=\frac{3}{40}\cdot\varphi(0.025)=\frac{3}{40}\cdot 0.3988\approx 0.03.$$

1.1.12. Интегральная теорема Лапласа

Производится n независимых испытаний, в каждом из которых может появиться или не появиться событие A. Вероятность наступления события A в каждом испытании постоянна и равна p $(0<p<1)$. Если число испытаний велико, то вероятность того, что событие A наступит не менее a раз и не более b раз, приближённо равна

$$p(a\le m\le b)\approx\Phi(x_2)-\Phi(x_1),$$

где m – число появлений событий A в n испытаниях;

$x_1=\dfrac{a-np}{\sqrt{npq}}$; $x_2=\dfrac{b-np}{\sqrt{npq}}$; $\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$ – функция Лапласа.

Свойства функции $\Phi(x)$:

1. Функция $\Phi(x)$ нечётная, то есть $\Phi(-x)=\Phi(x)$.
2. Для $x>5$ значение функции $\Phi(x)\approx 0{,}5$.

Значения функции $\Phi(x)$ приведены в приложении, табл. 2.

Пример 35. Известно, что при контроле бракуется 10 % деталей. Для контроля отобрано 500 изделий. Найти вероятность того, что число годных деталей окажется в пределах от 460 до 475.

Решение. По условию задачи

$$n = 500;\ p = 0{,}9;\ q = 0{,}1;\ a = 460;\ b = 475.$$

Подставляем

$$x_1 = \frac{a - np}{\sqrt{npq}} = \frac{460 - 500 \cdot 0{,}9}{\sqrt{500 \cdot 0{,}9 \cdot 0{,}1}} = 1{,}49;$$

$$x_2 = \frac{b - np}{\sqrt{npq}} = \frac{475 - 500 \cdot 0{,}9}{\sqrt{500 \cdot 0{,}9 \cdot 0{,}1}} = 3{,}73;$$

$$\Phi(x_1) = \Phi(1{,}49) = 0{,}4319;$$

$$\Phi(x_2) = \Phi(3{,}73) = 0{,}4999;$$

$$p(460 \le m \le 475) \approx \Phi(x_2) - \Phi(x_1) \approx 0{,}4999 - 0{,}4319 \approx 0{,}068.$$

Пример 36. Найти вероятность того, что событие *A* наступит ровно 70 раз в 243 испытаниях, если вероятность появления этого события в каждом испытании равна 0,25.

Решение: По условию, $n = 243;\ k = 70;\ p = 0.25;\ q = 0.75$. Так как $n = 243$ достаточно больше число, воспользуемся локальной теоремой Лапласа:

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

где $x = (k - np)/\sqrt{npq}$.

Найдем значение x:

$$x = \frac{k - np}{\sqrt{npq}} = \frac{70 - 243 \cdot 0.25}{\sqrt{243 \cdot 0.25 \cdot 0.75}} = \frac{9.25}{6.75} = 1.37$$

По таблице приложения 1 найдем $\varphi(1.37) = 0.1561$. Искомая вероятность

$$P_{243}(70) = 1/6.75 \cdot 0.1561 = 0.0231.$$

Пример 37. Игральный кубик подбрасывают 800 раз. Какая вероятность того, что количество очков, кратное трём, появится не менее 260 раз и не более 274 раз?

Решение. Для определения вероятности $P_{800}(260 \le m \le 274)$ применим интегральную теорему Муавра—Лапласа. Определим x_1, x_2:

$$x_1 = \frac{m_1 - np}{\sqrt{npq}} = \frac{260 - 800 \cdot \frac{1}{3}}{\frac{40}{3}} = \frac{20}{40} = -0.5;$$

$$x_2 = \frac{m_2 - np}{\sqrt{npq}} = \frac{274 - 800 \cdot \frac{1}{3}}{\frac{40}{3}} = \frac{22}{40} = 0.55.$$

Из прил. 1 находим

$$P_{800}(260 \le m \le 274) = \Phi(0.55) - \Phi(-0.5) = 0,2088 + 0,1915 = 0,04003.$$

1.1.13. Формула Пуассона

Производится n независимых испытаний, в каждом из которых может появиться или не появиться событие A. Вероятность наступления события A в каждом испытании постоянна и равна p. Если число испытаний велико, а вероятность p очень мала, то вероятность появления события A ровно m раз в n испытаниях приближённо определяется по формуле Пуассона:

$$p_n(m) \approx \frac{\lambda^m}{m!} \cdot e^{-\lambda},$$

где $\lambda = np$.

Формула Пуассона применяется для $\lambda \le 10$.

Пример 38. Вероятность появления события A в каждом из 250 независимых испытаний равна 0,008. Найти вероятность того, что событие A появится 3 раза.

Решение. По условию задачи имеем: $p = 0,008$; $n = 250$. Вычислим:

$$\lambda = np = 250 \cdot 0,008 = 2;$$

$$p_{250}(3) \approx \frac{2^3}{3!} e^{-2} \approx 0,18.$$

Пример 39. Учебник напечатан тиражом 90000 экземпляров. Вероятность неправильного брошюрования учебника равняется 0,0001. Найти вероятность того, что тираж содержит 5 бракованных учебников.

Решение. Брошюрование учебника можно рассматривать как испытание, которое вкладывается в схему Бернулли. Количество испытаний n большое, а вероятность каждого испытания p незначительная. Поэтому в этом случае целесообразно применить формул)’Пуассона.

В соответствии с условием задачи

$$n = 90000, \quad p = 0.0001$$

Итак, при $\lambda = np = 9$ из прил.3 получаем

$$P_{90000}(5) = \frac{9^5}{5!} \cdot e^{-9} \approx 0{,}0607.$$

1.1.14. Отклонение относительной частоты от постоянной вероятности в независимых испытаниях

Вероятность того, что в n независимых испытаниях, в каждом из которых вероятность появления события равна p $(0 < p < 1)$, абсолютная величина отклонения относительной частоты $\frac{m}{n}$ появления события от вероятности появления события не превышает положительного числа ε, приближённо равна удвоенной функции Лапласа $\Phi(x)$ при $x = \varepsilon \cdot \sqrt{\frac{n}{pq}}$:

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \approx 2 \cdot \Phi\left(\varepsilon \cdot \sqrt{\frac{n}{pq}}\right),$$

где $q = 1 - p$.

Пример 40. Вероятность того, что общая длина растений одного вида составляет 75–84 см, равна $0{,}6$. Какова вероятность того, что среди 300 таких растений относительная частота растений такой длины отклонится по абсолютной величине от вероятности появления растений такой длины не более чем на 0,05?

Решение. По условию:

$$n = 300;$$

$$\varepsilon = 0{,}05;$$

$$p = 0{,}6;$$

$$q = 1 - p = 0{,}4.$$

Тогда

$$P\left(\left|\frac{m}{n} - 0{,}6\right| \le 0{,}05\right) \approx 2 \cdot \Phi\left(0{,}05 \cdot \sqrt{\frac{300}{0{,}6 \cdot 0{,}4}}\right) \approx 2 \cdot \Phi(1{,}77) \approx 0{,}92$$

Пример 41. Производится 500 подбрасываний симметричной монеты. В каких пределах будет находиться отклонение частоты выпадения герба от 1/2 с вероятностью 0,95?

Решение. Используем формулу:

$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)$, где $p = 0.5,\ q = 1 - p = 0.5,\ \varepsilon = ?$ (отклонение), $n = 500$ - количество бросков монеты, $P = 0.95$- вероятность. Подставляем все:

$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = 2\Phi\left(\varepsilon\sqrt{\frac{m}{pq}}\right)$, $\Phi\left(\varepsilon\sqrt{\frac{500}{0.5 \cdot 0.5}}\right) = 0.475$, $\varepsilon\sqrt{\frac{500}{0.5 \cdot 0.5}} = 1.96$,

$\varepsilon = 1.96 \cdot 0.5 / \sqrt{500} \approx 0.044$.

Получаем пределы от $0.5 - 0.044 = 0.456$ до $0.5 + 0.044 = 0.544$.

Ответ. от 0,456 до 0,544.

1.2. Случайные величины

1.2.1. Закон распределения случайной величины

Случайной называют величину, которая принимает в результате испытания то или иное (но при этом только одно) возможное значение, заранее неизвестное, меняющееся от испытания к испытанию и зависящее от случайных обстоятельств.

В отличие от случайного события, являющегося качественной характеристикой случайного результата испытания, случайная величина характеризует результат испытания количественно.

Дискретной называют такую случайную величину, которая принимает конечное или бесконечное счётное множество значений.

Примером дискретной случайной величины могут являться: число дефектных деталей в партии, число заявок, число отказов элементов за определённое время и так далее.

Непрерывной называют такую случайную величину, которая может принимать любые значения из некоторого конечного или бесконечного интервала. Примером непрерывной случайной величины могут являться время безотказной работы отдельных элементов системы, погрешность измерения физических величин.

Случайные величины обычно обозначаются заглавными буквами конца латинского алфавита – $X, Y, Z, \ldots$, а их возможные значения – соответствующими малыми буквами – $x,\ y,\ z, \ldots$

Для задания случайной величины недостаточно перечислить все её возможные значения. Необходимо также знать, как часто могут появляться те или иные значения в результате испытаний при одних и тех же условиях, то есть нужно задать вероятности их появления.

Законом распределения случайной величины называется всякое соотношение между возможными значениями случайной величины и соответствующими вероятностями.

Две случайные величины называются *независимыми*, если распределение вероятности одной из них не зависит от того, какие возможные значения приняла другая величина. В противном случае случайные величины называются *зависимыми*.

Таблица, содержащая возможные значения случайной величины и соответствующие вероятности, называется *рядом распределения* случайной величины, и выглядит она следующим образом:

X	x_1	x_2	x_3	....	x_i	....
$p(X = x_i)$	p_1	p_2	p_3	....	p_i	....

Пример 42. Составить закон распределения числа появлений герба при трёх бросаниях монеты.

Решение. Случайная величина X – число появлений герба – может принимать значения: 0; 1; 2; 3. Соответствующие вероятности $p(X = 0)$, $p(X = 1)$, $p(X = 2)$, $p(X = 3)$ найдём по формуле Бернулли, используя тот факт, что вероятность появления герба в одном испытании равна $p = \frac{1}{2}$. Находим:

$$q = 1 - p = \frac{1}{2};$$

$$p(X = 0) = C_3^0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^3 = \frac{1}{8};$$

$$p(X = 1) = C_3^1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^2 = \frac{3}{8};$$

$$p(X = 2) = C_3^2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1 = \frac{3}{8};$$

$$p(X = 3) = C_3^3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^0 = \frac{1}{8}.$$

Следовательно, ряд распределения имеет вид:

X	0	1	2	3
$p(X = x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Пример 43. В студенческой группе организована лотерея. Разыгрывается две вещи стоимостью по 1000 руб. и одна стоимостью по 3000 руб. Составить закон распределения суммы чистого выигрыша для студента, который приобрёл один билет за 100 руб. Всего продано 50 билетов.

Решение. Интересующая нас случайная величина X может принимать три значения: -100 руб. (если студент не выиграет, а фактически проиграет 100 руб., уплаченные им за билет), 900 руб. и 2900 руб. (фактический выигрыш уменьшается на 100 руб. - на стоимость билета). Первому результату благоприятствуют 47 случаев из 50, второму - 2, а третьему - один. Поэтому их вероятности таковы: $P(X=-100)=47/50=0{,}94$, $P(X=900)=2/50=0{,}04$, $P(X=2900)=1/50=0{,}02$.

Закон распределения дискретной случайной величины X имеет вид

Сумма выигрыша	-100	900	2900
Вероятность	0,94	0,04	0,02

Пример 44. В урне 6 белых шаров и 4 чёрных шара. Из урны вынимают 3 шара. Число белых шаров среди вынутых шаров - дискретная случайная величина X. Составить соответствующий ей закон распределения.

Решение. Дискретная случайная величина X может принимать значения 0, 1, 2, 3. Соответствующие им вероятности проще всего вычислить по **правилу умножения вероятностей.** Получаем следующий закон распределения дискретной случайной величины:

Значение	0	1	2	3
Вероятность	1/30	3/10	1/2	1/6

1.2.2. Функция распределения

Функция распределения является наиболее общей формой задания закона распределения случайной величины. Она используется как для дискретных, так и для непрерывных случайных величин. Обозначается $F(x)$.

Функция распределения задаёт вероятность того, что случайная величина X примет значение, меньше фиксированного действительного числа x, то есть $F(x) = p(X < x)$.

Свойства функции распределения.

1. Функция распределения $F(x)$ есть неотрицательная функция, заключённая в промежутке: $0 \le F(x) \le 1$.

2. Функция распределения $F(x)$ случайной величины есть неубывающая функция, то есть из $x_2 > x_1$ следует $F(x_2) > F(x_1)$.

3. На минус бесконечности функция $F(x)$ распределения равна нулю, а на плюс бесконечности функция $F(x)$ распределения равна единице, то есть $F(-\infty) = 0;\ \ F(\infty) = 1$.

Следует отметить, что в то время как каждая случайная величина однозначно задаёт функцию распределения, одну и ту же функцию распределения могут иметь различные случайные величины. Отсюда следует второе определение случайной величины.

Случайной величиной называется переменная величина, значения которой зависят от случая и для которой задана функция распределения вероятностей.

Теорема 5. Вероятность попадания случайной величины в интервал $[\alpha, \beta]$ равна разности значений функции распределения $F(x)$ на концах этого интервала:

$$p(\alpha \le X \le \beta) = F(\beta) - F(\alpha).$$

Следствие. Вероятность любого отдельного значения непрерывной случайной величины равна нулю.

Пример 45. Составить функцию распределения для ряда распределения из примера 20.

Решение. Случайная величина X – число появлений герба. Из примера 19 имеем:

$$p(X = 0) = \frac{1}{8};\ \ p(X = 1) = \frac{3}{8};\ \ p(X = 2) = \frac{3}{8};\ \ p(X = 3) = \frac{1}{8}.$$

По определению функции распределения $F(x)$ можем записать:

$$F(x)=0 \text{ для } x<0;$$

$$F(x)=\frac{1}{8} \text{ для } 0\le x<1;$$

$$F(x)=\frac{1}{8}+\frac{3}{8}=\frac{4}{8} \text{ для } 1\le x<2;$$

$$F(x)=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}=\frac{7}{8} \text{ для } 2\le x<3;$$

$$F(x)=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=1 \text{ для } 3\le x.$$

Классически функция распределения $F(x)$ записывается следующим образом:

$$F(x)=\begin{cases} 0 & x<0; \\ \frac{1}{8} & 0\le x<1; \\ \frac{4}{8} & 1\le x<2; \\ \frac{7}{8} & 2\le x<3; \\ 1 & 3\le x. \end{cases}$$

Пример 46. Составить закон распределения дискретной случайной величины - числа попаданий в цель при четырёх выстрелах, если вероятность попадания при одном выстреле равна 0,1.

Решение. Дискретная случайная величина *X* может принимать пять различных значений: 1, 2, 3, 4, 5. Соответствующие им вероятности найдём по **формуле Бернулли** $P_{m,n}=C_n^m p^m q^{n-m}$. При

$$\begin{aligned} &n=4, \\ &p=1.1 \\ &q=1-p=0.9 \\ &m=0,1,2,3,4 \end{aligned}$$

получаем

$$P_{0.4} = 0.9^4 = 0.6561;$$
$$P_{1.4} = 4 \cdot 0.1 \cdot 0.9^3 = 0.2961;$$
$$P_{2.4} = 6 \cdot 0.1^2 \cdot 0.9^2 = 0.0486;$$
$$P_{3.4} = 4 \cdot 0.1^3 \cdot 0.9 = 0.0036;$$
$$P_{4.4} = 0.1^4 = 0.0001.$$

Следовательно, закон распределения дискретной случайной величины X имеет вид

Число попаданий	0	1	2	3	4
Вероятность	0,6561	0,2916	0,0486	0,0036	0,0001

Пример 47. Построить функцию распределения случайной величины x

x_1	-2	0	3	7
p_1	0.4	0.1	0.3	0.2

Найти вероятности того, что случайная величина примет значение из следующих промежутков:

$$P(-1 < X < 5),\ P(4 < X \le 10),\ P(X \le 2),\ P(3 \le X \le 7),\ P(X > 7),\ P(|\,X - M(X)\,| < \sigma(X))$$

Решение. Рассмотрим формальный алгоритм построения функции распределения

Сначала берём первое значение $x_1 = -2$ и составляем *нестрогое* неравенство $x_1 \le -2$. На этом промежутке $F(x) = 0$.

На промежутке $-2 < x \le 0$ (между x_1 и x_2): $F(x) = 0,4$.

На промежутке $0 < x \le 3$(между x_2 и x_3: $F(x) = 0,4 + 0,1 = 0,5$.

На промежутке $3 < x \le 7$ (между x_3 и x_4): И конец если x строго больше самого последнего значения $x_4 = 7$, то : $F(x) = 0,4 + 0,1 + 0,3 + 0,2 = 0,8 + 0,2 = 1$.

Легко заметить, что с увеличением «икс» идет накопление (суммирование) вероятностей, и поэтому функцию $P(x)$ также называют *интегральной* функцией распределения. В практических задачах проведенные выше действия обычно выполняют в уме, а результат сразу записывают под единую скобку:

1.2.3. Плотность распределения

Непрерывную случайную величину можно задать не только интегральной функцией распределения, но и дифференциальной функцией. Рассмотрим эту форму задания распределения случайной величины. Пусть задана непрерывная случайная величина X с функцией распределения $F(x)$. Тогда, если существует

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

то функция $f(x) = F'(x)$ называется дифференциальной функцией распределения или плотностью распределения.

Используя методы интегрального исчисления, можно записать формулу для нахождения интегральной функции распределения по плотности:

$$F(x) = \int_{-\infty}^{x} f(x) dx.$$

Свойства плотности распределения:

1. Плотность распределения больше либо равна нулю для любого значения аргумента, то есть $f(x) \geq 0$.

2. Так как интегральная функция распределения неубывающая, следовательно, её производная неотрицательная.

3. Вероятность попадания случайной величины в заданный интервал $[\alpha, \beta]$ находится по формуле

$$p(\alpha \leq x < \beta) = \int_{\alpha}^{\beta} f(x) dx.$$

4. Условие нормировки: интеграл в бесконечных пределах от плотности распределения равен единице:

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

Пример 48. Случайная величина X подчинена закону распределения с плотностью

$$f(x)=\begin{cases}0 & x<0;\\ c\cdot\sin x & 0\le x\le\pi;\\ 0 & \pi<x.\end{cases}$$

Определить коэффициент c. Найти функцию распределения и вероятность попадания случайной величины на интервал от 0 до $\frac{\pi}{2}$.

Решение. Согласно условию нормировки имеем:

$$\int\limits_{-\infty}^{\infty} f(x)dx=1,$$

отсюда можно записать:

$$\int\limits_{-\infty}^{\infty} f(x)dx=\int\limits_{-\infty}^{0} f(x)dx+\int\limits_{0}^{\pi} f(x)dx+\int\limits_{\pi}^{\infty} f(x)dx=$$

$$=\int\limits_{-\infty}^{0} 0dx+\int\limits_{0}^{\pi} c\cdot\sin x dx+\int\limits_{\pi}^{\infty} 0dx=$$

$$=c\cdot\int\limits_{0}^{\pi}\sin x dx=-c\cdot\cos x\Big|_{0}^{\pi}=-c(-1-1)=2c,$$

тогда $2\cdot c=1;\quad c=\frac{1}{2}.$

Вероятность попадания в интервал найдём по формуле

$$p\left(0\le x\le\frac{\pi}{2}\right)=\int\limits_{0}^{\frac{\pi}{2}}\frac{1}{2}\sin x dx=\frac{1}{2}\cos x\Big|_{0}^{\frac{\pi}{2}}=-\frac{1}{2}(0-1)=\frac{1}{2}.$$

Функцию распределения $F(x)$ найдём по формуле

$$F(x)=\int\limits_{-\infty}^{x} f(x)dx.$$

Для $x<0$ $F(x)=\int\limits_{-\infty}^{x}0dx=0$; для $0\le x<\pi$

$$F(x)=\int\limits_{-\infty}^{0}0dx+\int\limits_{0}^{x}\frac{1}{2}\sin xdx=0-\frac{1}{2}\cos x\Big|_{0}^{x}=\frac{1}{2}(1-\cos x);$$ для $\pi\le x$

$$F(x)=\int\limits_{-\infty}^{0}0dx+\int\limits_{0}^{\pi}\frac{1}{2}\sin xdx+\int\limits_{\pi}^{\infty}0dx=0-\frac{1}{2}\cos x\Big|_{0}^{\pi}+0=-\frac{1}{2}(-1-1)=1.$$

Таким образом получаем:

$$F(x)=\begin{cases}0 & x<0;\\ \frac{1}{2}(1-\cos x) & 0\le x\le\pi;\\ 1 & x\ge\pi.\end{cases}$$

Пример 49. Дана функция распределения непрерывной случайной величины X

$$F(x)=\begin{cases}0 & при\ x\le 0\\ \sin x & при\ 0<x\le\pi/2\\ 1 & при\ x>\pi/2\end{cases}$$

Найти плотность распределения $f(x)$.

Решение. Плотность распределения равна первой производной от функции распределения:

$$f(x)=F'(x)=\begin{cases}0\ при\ x<0,\\ \cos x\ при\ 0<x\le\pi/2\\ 0\ при\ x>\pi/2\end{cases}$$

Заметим, что при $x=0$ производная $F'(x)$ не существует.

Пример 50. Непрерывная случайная величина X задана плотностью распределения $f(x)=(3/2)\sin 3x$ в интервале $(0,\pi/3)$; вне этого интервала $f(x)=0$. Найти вероятность того, что X примет значение, принадлежащее интервалу $(\pi/6,\pi/4)$.

Решение. Воспользуемся формулой $P(a<X<b)=\int\limits_{a}^{b}f(x)dx.$ По условию, $a=\pi/6, b=\pi/4,$ $f(x)=(3/2)\sin 3x.$ Следовательно, искомая вероятность

$$P(\pi/6 < X < \pi/4) = (3/2)\int_{\pi/6}^{\pi/4} \sin 3x dx = \sqrt{2/4}.$$

Пример 51. Задана плотность $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Нормально распределенной случайной величины X. Найти плотность распределения $g(y)$. Случайной величины $Y = X^2$.

Решение. Так как в интервале $(-\infty;+\infty)$ функция $y = x^2$ не монотонна, то разобьем этот интервал на интервалы $(-\infty;0)$ и $(0;+\infty)$, в которых она монотонна. В интервале $(-\infty;0)$ обратная функция $\psi_1(y) = -\sqrt{y}$, в интервале $(0;+\infty)$ $\psi_2(y) = \sqrt{y}$,

$$|\psi_1(y)| = \frac{1}{2\sqrt{y}},\ f(\psi_1(y)) = \frac{1}{2\sqrt{y}} e^{-\frac{y}{2}},\ f(\psi_2(y)) = \frac{1}{2\sqrt{\pi}} e^{-\frac{y}{2}}.$$

Искомую плотность распределения находим из равенства

$$g(y) = f(\psi_1(y)) \cdot |\psi_1(y)| + f(\psi_2(y)) \cdot |\psi_2(y)|,$$

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}} e^{-\frac{y}{2}}.$$

Так как $y = x^2$, причем $-\infty < x < +\infty$, то $0 < y < \infty$. Таким образом, в интервале $(0;+\infty)$ искомая плотность распределения $g(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}$, вне этого интервала $g(y) = 0$.

Ответ. $g(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}$ при, $(0;+\infty)$, $g(y) = 0$ при $y \in (-\infty;0)$.

1.2.4. Математическое ожидание

Математическое ожидание определяет положение случайной величины на числовой оси, показывая центр распределения (некоторое среднее значение, около которого сосредоточены все возможные значения случайной величины).

Математическое ожидание дискретной случайной величины – это число, равное сумме произведений всех возможных значений на их вероятности, то есть

$$M(x)=\sum_{k=1}^{n} x_k \cdot p_k .$$

Математическое ожидание непрерывной случайной величины X находится по формуле

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

Свойства математического ожидания:

1. Математическое ожидание постоянной равно самой постоянной, то есть $M(c)=c$, где $c-\text{const}$.

2. Постоянный множитель выносится за знак математического ожидания, то есть $M(c \cdot x)=c \cdot M(x)$.

3. Математическое ожидание суммы двух независимых случайных величин равно сумме их математических ожиданий, то есть

$$M(x+y)=M(x)+M(y).$$

4. Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий, то есть

$$M(x \cdot y)=M(x) \cdot M(y).$$

5. Математическое ожидание отклонения случайной величины от её математического ожидания равно нулю, то есть $M(x-M(x))=0$.

Пример 52. Найти математическое ожидание для случайной величины, заданной функцией распределения:

$$F(x)=\begin{cases} 0 \quad x<0; \\ 1-\cos x \quad 0 \le x < \dfrac{\pi}{2}; \\ 1 \quad x>\dfrac{\pi}{2}. \end{cases}$$

Решение. Найдём плотность распределения по формуле $f(x)=F'(x)$:

$$f(x) = F'(x) = \begin{cases} 0 & x < 0; \\ \sin x & 0 \le x < \dfrac{\pi}{2}; \\ 0 & x > \dfrac{\pi}{2}. \end{cases}$$

Математическое ожидание непрерывной случайной величины находим по формуле

$$M(x) = \int\limits_{-\infty}^{\infty} x \cdot f(x)dx = \int\limits_{0}^{\frac{\pi}{2}} x \cdot \sin x dx = \left\langle \begin{matrix} u = x & dv = \sin x dx \\ du = dx & v = -\cos x \end{matrix} \right\rangle =$$

$$= -x\cos x \Bigg|_{0}^{\frac{\pi}{2}} + \int\limits_{0}^{\frac{\pi}{2}} \cos x dx = \sin x \Bigg|_{0}^{\frac{\pi}{2}} = 1.$$

Пример 53. Найти математическое ожидание случайной величины, заданной рядом распределения

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

Решение. Для дискретной случайной величины используем формулу

$$M(x) = \sum_{k=1}^{n} x_k \cdot p_k = \sum_{k=1}^{4} x_k \cdot p_k = 0 \cdot 0{,}1 + 1 \cdot 0{,}3 + 2 \cdot 0{,}4 + 3 \cdot 0{,}2 = 1{,}7.$$

Модой M_0 *дискретной случайной величины* называется её значение, имеющее наибольшую вероятность.

Модой M_0 *непрерывной случайной величины* называется такое её значение, при котором плотность распределения имеет максимум.

Геометрически моду можно интерпретировать как абсциссу точки максимума кривой распределения. Бывают *двухмодульные* и *многомодульные* распределения. Встречаются распределения, которые имеют минимум, но не имеют максимума. Такие распределения называются *антимодальными*.

Медианой случайной величины M_e называют такое её значение, для которого справедливо равенство $p(X < M_e) = p(X > M_e)$, то есть равновероятно, что случайная величина окажется больше или меньше медианы.

С геометрической точки зрения, медиана – абсцисса точки, в которой площадь, ограниченная кривой распределения, делится пополам. Так как вся площадь, ограниченная кривой распределения и осью абсцисс, равна единице, то функция распределения в точке, соответствующей медиане, равна 0,5, то есть

$$F(M_e) = p(X < M_e).$$

1.2.5. Дисперсия и среднее квадратическое отклонение

Дисперсией случайной величины X называют математическое ожидание квадрата отклонения случайной величины от её математического ожидания $D(x) = M(x - M(x))^2$.

Дисперсия характеризует меру рассеяния случайной величины вокруг математического ожидания. Недостатком дисперсии является то, что она имеет размерность квадрата случайной величины и её неудобно использовать для характеристики разброса.

Этих недостатков лишено *среднее квадратическое отклонение* случайной величины, которое представляет собой квадратный корень из дисперсии $\sigma(x) = \sqrt{D(x)}$.

Основные свойства дисперсии:

1. Дисперсия постоянной величины равна нулю, $D(c) = 0$, где $c - \text{const}$.

2. Постоянный множитель случайной величины можно выносить за знак дисперсии, предварительно возведя его в квадрат $D(c \cdot x) = c^2 \cdot D(x)$.

3. Дисперсия суммы двух независимых случайных величин равна сумме дисперсий этих величин: $D(x + y) = D(x) + D(y)$.

4. Дисперсия разности двух независимых случайных величин равна сумме их дисперсий, $D(x-y)=D(x)+D(y)$.

5. Дисперсия случайной величины X равна разности между математическим ожиданием квадрата случайной величины X и квадратом её математического ожидания, $D(x)=M\left(x^2\right)-M^2(x)$.

Пример 54. Найти дисперсию для распределения непрерывной случайной величины, заданной в примере 23.

Решение. В примере 23 было найдено математическое ожидание $M(x)=1$. Для нахождения дисперсии используем её 5 свойство. Вычислим:

$$M\left(x^2\right)=\int_{-\infty}^{\infty} x^2\cdot f(x)dx=\int_{0}^{\frac{\pi}{2}} x^2\cdot\sin x dx=\left\langle \begin{matrix} u=x^2 & dv=\sin x dx \\ du=2xdx & v=-\cos x \end{matrix}\right\rangle=$$

$$=-x^2\cos x\Big|_{0}^{\frac{\pi}{2}}+\int_{0}^{\frac{\pi}{2}} 2x\cdot\cos x dx=\left\langle \begin{matrix} u=x & dv=\sin x dx \\ du=dx & v=-\cos x \end{matrix}\right\rangle=$$

$$=0+2x\cdot\sin x\Big|_{0}^{\frac{\pi}{2}}-\int_{0}^{\frac{\pi}{2}} 2\cdot\sin x dx=\pi+2\cdot\cos x\Big|_{0}^{\frac{\pi}{2}}=\pi-2.$$

Тогда дисперсия равна

$$D(x)=M\left(x^2\right)-M^2(x)=\pi-2-1=\pi-3.$$

Пример 55. Найти дисперсию и среднее квадратическое отклонение для дискретной случайной величины, имеющей следующий ряд распределения:

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

Решение. Математическое ожидание для этого распределения найдено в примере 24 и равно $M(x)=1{,}7$. Для нахождения дисперсии вычислим:

$$M\left(x^2\right)=\sum_{k=1}^{4} x_k^2\cdot p_k=0^2\cdot 0{,}1+1^2\cdot 0{,}3+2^2\cdot 0{,}4+3^2\cdot 0{,}2=3{,}7.$$

Тогда дисперсия равна $D(x) = M(x^2) - M^2(x) = 3{,}7 - 1{,}7^2 = 0{,}81$ Среднее квадратическое отклонение найдём по формуле $\sigma(x) = \sqrt{D(x)} = \sqrt{0{,}81} = 0{,}9$.

Пример 56. Дискретная случайная величина X имеет только два возможных значения: x_1 и x_2, причём $x_1 < x_2$. Вероятность того, что X примет значение x_1, равна 0,5. Найти закон распределения X, зная математическое ожидание $M(x) = 4$ и $D(x) = 1$.

Решение. По условию нормировки имеем $p_1 + p_2 = 1$, отсюда $p_2 = 1 - p_1 = 1 - 0{,}5 = 0{,}5.$

Так как $M(x) = x_1 \cdot p_1 + x_2 \cdot p_2 = 0{,}5 \cdot x_1 + 0{,}5 \cdot x_2 = 4$, можем записат $x_1 + x_2 = 8$. Выразим одну неизвестную через другую $x_2 = 8 - x_1$. По пятому свойству дисперсии $D(x) = M(x^2) - M^2(x)$, Отсюда можем записать

$$M(x^2) = D(x) + M^2(x) = 1 + 4^2 = 17,$$

тогда

$$M(x^2) = x_1^2 \cdot p_1 + x_2^2 \cdot p_2 = x_1^2 \cdot 0{,}5 + x_2^2 \cdot 0{,}5 = 17.$$

Отсюда $x_1^2 + x_2^2 = 34$. Подставим $x_2 = 8 - x_1$ в полученное уравнение $x_1^2 + x_2^2 = 34$, получим $x_1^2 + (8 - x_1)^2 = 34$ преобразуем полученное соотношение $x_1^2 - 8x_1 + 15 = 0$. Решая полученное квадратное уравнение, находим корни $x_{1,1} = 5$ и $x_{1,2} = 3$. Тогда $x_{2,1} = 3$ и $x_{2,2} = 5$. По условию задачи $x_1 < x_2$, следовательно: $x_1 = 3$, а $x_2 = 5$ и искомый закон распределения имеет вид:

x_i	3	5
p_i	0,5	0,5

1.2.6. Начальные и центральные моменты

Начальным моментом k*-го порядка* случайной величины называют математическое ожидание величины X^k, то есть $V_k = M\left(X^k\right)$.

Начальный момент для дискретной случайной величины

$$V_k = \sum_{i=1}^{n} x_i^k \cdot p_i .$$

Начальный момент для непрерывной случайной величины

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

Центральным моментом k*-го порядка* случайной величины называют математическое ожидание k*-й* степени отклонения случайной величины от его математического ожидания, то есть $\mu_k = M^k(x - M(x))$.

Центральный момент для дискретной случайной величины находится по формуле

$$\mu_k = \sum_{i=1}^{n} (x - M(x))^k \cdot p_i .$$

Центральный момент для непрерывной случайной величины находится по формуле

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

Соотношения между начальными и центральными моментами

$$\mu_0 = 1:\ \mu_1 = 0;\ \mu_2 = V_2 - V_1^2;\ \mu_3 = V_3 - 3V_2 \cdot V_1 + 2V_1^3;$$

$$\mu_4 = V_4 - 4V_3 \cdot V_1 + 6V_2 \cdot V_1^2 - 3V_1^4.$$

Начальный момент первого порядка представляет собой математическое ожидание $V_1 = M(x)$, а центральный момент второго порядка – дисперсию случайной величины $\mu_2 = D(x)$.

Нормированный центральный момент третьего порядка служит характеристикой скошенности или асимметрии распределения (*коэффициент асимметрии*): $A = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\sqrt{\mu_2^3}}$.

Нормированный центральный момент четвёртого порядка служит характеристикой островершинности или плосковершинности распределения (*эксцесс*), $E = \frac{\mu_4}{\sigma^4} - 3 = \frac{\mu_4}{\mu_2^2} - 3$.

Пример 57. Случайная величина X задана плотностью распределения

$$f(x) = \begin{cases} 0 & x < 0; \\ a \cdot x^2 & 0 \le x \le 2; \\ 0 & x \ge 2. \end{cases}$$

Найти коэффициент a, математическое ожидание, дисперсию, коэффициенты асимметрии и эксцесса.

Решение. Площадь, ограниченная плотностью распределения, численно равна $\int\limits_{-\infty}^{\infty} f(x)dx = \int\limits_{0}^{2} a \cdot x^2 dx = a\frac{x^3}{3}\Big|_0^2 = \frac{8a}{3}$. Учитывая, что эта площадь должна быть равна единице, получим $a = \frac{3}{8}$.

Найдём начальные моменты:

$$V_1 = \int\limits_{-\infty}^{\infty} x \cdot f(x)dx = \int\limits_{0}^{2} \frac{3}{8} \cdot x^3 dx = \frac{3}{8} \cdot \frac{x^4}{4}\Big|_0^2 = \frac{3}{2};$$

$$V_2 = \int\limits_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int\limits_{0}^{2} \frac{3}{8} \cdot x^4 dx = \frac{3}{8} \cdot \frac{x^5}{5}\Big|_0^2 = \frac{12}{5};$$

$$V_3 = \int\limits_{-\infty}^{\infty} x^3 \cdot f(x)dx = \int\limits_{0}^{2} \frac{3}{8} \cdot x^5 dx = \frac{3}{8} \cdot \frac{x^6}{6}\Big|_0^2 = 4;$$

$$V_4 = \int_{-\infty}^{\infty} x^4 \cdot f(x)dx = \int_0^2 \frac{3}{8} \cdot x^6 dx = \frac{3}{8} \cdot \frac{x^7}{7}\bigg|_0^2 = \frac{48}{7}.$$

Найдём центральные моменты:

$$\mu_2 = V_2 - V_1^2 = \frac{12}{5} - \left(\frac{3}{2}\right)^2 = \frac{3}{20};$$

$$\mu_3 = V_3 - 3V_2 \cdot V_1 + 2V_1^3 = 4 - 3 \cdot \frac{3}{2} \cdot \frac{12}{5} + 2 \cdot \left(\frac{3}{2}\right)^3 = -\frac{1}{20};$$

$$\mu_4 = V_4 - 4V_3 \cdot V_1 + 6V_2 \cdot V_1^2 - 3V_1^4 =$$

$$= \frac{48}{7} - 4 \cdot 4 \cdot \frac{3}{2} + 6 \cdot \frac{12}{5} \cdot \left(\frac{3}{2}\right)^2 - 3 \cdot \left(\frac{3}{2}\right)^4 = \frac{39}{560}.$$

Вычислим остальные характеристики:

математическое ожидание $M(x) = V_1 = \frac{3}{2}$;

дисперсия равна $D(x) = \mu_2 = \frac{3}{20}$;

асимметрия $A = \frac{\mu_3}{\sqrt{\mu_2^3}} = -\frac{1}{20} \cdot \frac{20}{3} \cdot \sqrt{\frac{20}{3}} = -\frac{\sqrt{20}}{3 \cdot \sqrt{3}} \approx 0{,}86$;

эксцесс $E = \frac{\mu_4}{\mu_2^2} - 3 = \frac{39}{560} \cdot \frac{400}{9} - 3 = \frac{2}{21}$.

1.2.7. Равномерное распределение

Непрерывная случайная величина X имеет *равномерное распределение* на интервале $[a, b]$, если на этом интервале плотность распределения постоянна, а вне его равна нулю, то есть

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

Найдём значение постоянной c. Так как площадь, ограниченная кривой распределения, равна единице и все значения случайной величины принадлежат интервалу $[a, b]$, то должно выполняться равенство

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

или

$$\int_a^b cdx = c \cdot x \Big|_a^b = c \cdot (b - a) = 1; \qquad c = \frac{1}{b-a}.$$

Следовательно, плотность распределения можно записать в виде:

$$f(x) = \begin{cases} 0 \quad x < a; \\ \dfrac{1}{b-a} \quad a \le x \le b; \\ 0 \quad x > b. \end{cases}$$

Интегральная функция распределения имеет вид:

$$F(x) = \begin{cases} 0 \quad x < a; \\ \dfrac{x-a}{b-a} \quad a \le x \le b; \\ 1 \quad x > b. \end{cases}$$

Определим основные числовые характеристики случайной величины, имеющей равномерное распределение.

Математическое ожидание

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

В силу симметричности распределения медиана совпадает с математическим ожиданием, то есть

$$M_e = \frac{b+a}{2}.$$

Моды равномерное распределение не имеет. Дисперсия равномерного распределения равна

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

В силу симметричности коэффициент асимметрии равен нулю. Коэффициент эксцесса равен $E=-1{,}2$.

Пример 58. Цена деления амперметра равна $0{,}1\,A$. Показания округляются до ближайшего целого деления. Найти вероятность того, что при отсчёте будет сделана ошибка, превышающая $0{,}02\,A$. Вычислить математическое ожидание и дисперсию.

Решение. Ошибку округления можно рассматривать как случайную величину X, которая распределена равномерно в интервале между двумя соседними целыми делениями. Плотность равномерного распределения $f(x)=\dfrac{1}{(b-a)}$, где $(b-a)$ – длина интервала, в котором заключены возможные значения, вне этого интервала $f(x)=0$. В рассматриваемом примере длина интервала, в котором заключены возможные значения X, равна $0{,}1$, поэтому $f(x)=\dfrac{1}{0{,}1}=10$. Ошибка отсчёта превысит $0{,}02$, если она будет заключена в интервале $(0{,}02;\,0{,}08)$. По формуле

$$p(\alpha<x<\beta)=\int\limits_{\alpha}^{\beta}f(x)dx$$

получим

$$p(0{,}02<x<0{,}08)=\int\limits_{0{,}02}^{0{,}08}10dx=10x\Big|_{0{,}02}^{0{,}08}=10\cdot(0{,}08-0{,}02)=0{,}6.$$

Математическое ожидание равно

$$M(x)=\frac{b+a}{2}=\frac{0{,}1+0}{2}=0{,}05.$$

Дисперсия равна

$$D(x)=\frac{(b-a)^2}{12}=\frac{(0{,}1-0)^2}{12}=\frac{1}{1200}.$$

Пример 59. Погрешность измерения прибора распределена равномерно на интервале [5;10]. Найти вероятность того, что погрешность прибора не превышает 7 ед. Вычислить погрешность измерения, вероятность которой равна 0,95. Вычислить вероятность того, что погрешность измерения будет находиться в интервале(6;8) ед.

Решение. Так как имеем равномерно распределенную случайную величину, то вероятность того, что погрешность не превысит 7 ед., равна

$$F(7)=\frac{7-5}{10-5}=0{,}4.$$

Для вычисления значения погрешности измерения воспользуемся условием:

$$F(x)=\frac{x-a}{b-a}=1-0{,}95=0{,}05,$$

получаем $x_{0,95}=(10-5)\cdot 0{,}05+5=5{,}25.$

Вероятность того, что погрешность измерения будет находится в интервале (6:8), находим из условия:

$$P(6\le x\le 8)=F(8)-F(6)=\frac{8-5}{10-5}-\frac{6-5}{10-5}=0{,}4.$$

1.2.8. Нормальное распределение

Нормальное распределение – наиболее часто встречающийся вид распределения. С ним приходится встречаться при анализе погрешностей измерений, контроле технологических процессов и режимов, а также при анализе и прогнозировании различных явлений в биологии, медицине и других областях знаний.

Термин «нормальное распределение» применяется в условном смысле как общепринятый в литературе, хотя и не совсем удачный. Так утверждение, что какой-то признак подчиняется нормальному закону распределения, вовсе не означает наличие каких-либо незыблемых норм, якобы лежащих в основе явления, отражением которого является рассматриваемый признак, а подчинение другим законам распределения не означает какую-то анормальность данного явления.

Главная особенность нормального распределения состоит в том, что оно является предельным, к которому приближаются другие распределения. Нормальное распределение впервые открыто Муавром в 1733 году. Нормальному закону подчиняются только непрерывные случайные величины. Плотность нормального закона распределения

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

Математическое ожидание для этого распределения равно $M(x)=a$. Дисперсия равна $D(x)=\sigma^2$.

Основные свойства нормального распределения:

1. Функция плотности распределения определена на всей числовой оси Ox, то есть каждому значению x соответствует вполне определённое значение функции.

2. При всех значениях x (как положительных, так и отрицательных) функция плотности принимает положительные значения, то есть нормальная кривая расположена над осью Ox.

3. Предел функции плотности при неограниченном возрастании x равен нулю, $\lim\limits_{x\to\pm\infty} f(x)=0$.

4. Функция плотности нормального распределения в точке $x=a$ имеет максимум $f(a)=\frac{1}{\sigma\sqrt{2\pi}}$.

5. График функции плотности $y=f(x)$ симметричен относительно прямой $x=a$.

6. Кривая распределения имеет две точки перегиба с координатами $\left(a-\sigma;\ \dfrac{1}{\sigma\sqrt{2\pi e}}\right)$ и $\left(a+\sigma;\ \dfrac{1}{\sigma\sqrt{2\pi e}}\right)$.

7. Мода и медиана нормального распределения совпадают с математическим ожиданием a.

8. Форма нормальной кривой не изменяется при изменении параметра a.

9. Коэффициенты асимметрии и эксцесса нормального распределения равны нулю.

Очевидна важность вычисления этих коэффициентов для эмпирических рядов распределения, так как они характеризуют скошенность и крутость данного ряда по сравнению с нормальным.

Вероятность попадания в интервал $(\beta;\gamma)$ находится по формуле

$$p(\beta \le x \le \gamma) = \Phi\left(\frac{\gamma-\alpha}{\sigma}\right) - \Phi\left(\frac{\beta-\alpha}{\sigma}\right),$$

где $\Phi(x) = \dfrac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$ – нечётная табулированная функция.

Определим вероятность того, что нормально распределённая случайная величина отклоняется от своего математического ожидания на величину, меньшую ε, то есть найдём вероятность осуществления неравенства $|x - M(x)| < \varepsilon$ или вероятность двойного неравенства $M(x) - \varepsilon < x < M(x) + \varepsilon$. Подставляя в формулу, получим

$$p(a-\varepsilon \le x \le a+\varepsilon) = \Phi\left(\frac{a+\varepsilon-\alpha}{\sigma}\right) - \Phi\left(\frac{a-\varepsilon-\alpha}{\sigma}\right) =$$

$$= \Phi\left(\frac{\varepsilon}{\sigma}\right) - \Phi\left(-\frac{\varepsilon}{\sigma}\right) = \Phi\left(\frac{\varepsilon}{\sigma}\right) + \Phi\left(\frac{\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

Выразив отклонение случайной величины X в долях среднего квадратического отклонения, то есть положив $\varepsilon = t\cdot\sigma$ в последнем равенстве, получим $p\left(\left|x-M(x)\right|<t\cdot\sigma\right)=2\Phi(t)$. Тогда при различных значениях t получим

$$t=1\ \ p\left(\left|x-M(x)\right|<\sigma\right)=2\Phi(1)\approx 0{,}6827;$$

$$t=2\ \ p\left(\left|x-M(x)\right|<2\sigma\right)=2\Phi(2)\approx 0{,}9545;$$

$$t=3\ \ p\left(\left|x-M(x)\right|<3\sigma\right)=2\Phi(3)\approx 0{,}9973.$$

Из последнего соотношения следует, что практически рассеяние нормально распределённой случайной величины заключено на участке $(a-3\sigma;\, a+3\sigma)$. Вероятность того, что случайная величина не попадёт на этот участок, очень мала, а именно равна $0{,}0027$, то есть это событие может произойти лишь в трёх случаях из 1000. Такие события можно считать практически невозможными. На приведённых рассуждениях основано правило трёх сигм, которое формулируется следующим образом: *если случайная величина имеет нормальное распределение, то отклонение этой величины от математического ожидания по абсолютной величине не превосходит утроенного среднего квадратического отклонения*.

Пример 60. Деталь, изготовленная автоматом, считается годной, если отклонение её контролируемого размера от проектного не превышает 10 мм. Случайные отклонения контролируемого размера от проектного подчинены нормальному закону распределения со средним квадратическим отклонением $\sigma=5$ мм и математическим ожиданием $a=0$. Сколько процентов годных деталей изготавливает автомат?

Решение. Рассмотрим случайную величину X – отклонение размера от проектного. Деталь будет признана годной, если случайная величина принадлежит интервалу $[10;10]$. Вероятность изготовления годной детали найдём по формуле $p\left(|x|<10\right)=2\Phi\left(\dfrac{10}{5}\right)=2\Phi(2)\approx 0{,}9544$. Следовательно, процент годных деталей, изготавливаемых автоматом, равен 95,44 %.

Пример 61. Определить закон распределения случайной величины X, если ее плотность распределения вероятностей задана функцией

$$p(x)=\frac{1}{\sigma\cdot\sqrt{2\pi}}e^{-\frac{(x-1)^2}{72}}.$$

Найти математическое ожидание, дисперсию и функцию распределения случайной величины X.

Решение. Сравнивая данную функцию $P(X)$ с функцией плотности вероятности для случайной величины, распределенной по нормальному закону, заключаем, что случайная величина X распределена по нормальному закону с параметрами $a=1$ и $\sigma=6$. Тогда

$$M(X)=1,\ \sigma(X)=6,\ D(X)=36.$$

Функция распределения случайной величины *X* имеет вид

$$F(x)=\frac{1}{2}+\frac{1}{2}\Phi(\frac{x-1}{6}).$$

Пример 62. Текущая цена акции может быть смоделирована с помощью нормального закона распределения с математическим ожиданием 15 ден. ед. и средним квадратическим отклонением 0,2 ден. ед.

Найти вероятность того, что цена акции: а) не выше 15.3 ден ед.; б) не ниже 15,4 ден ед., в) от 14.9 до 15,3 ден. ед. С помощью «правила трех сигм» найти границы, в которых будет находиться текущая цена акции.

Решение. Так как $a=15$ и $\sigma=0,2$, то

$$P(X\le 15,3)=F(15,3)=\frac{1}{2}+\frac{1}{2}\Phi(\frac{15,3-15}{0,2})=\frac{1}{2}+\frac{1}{2}\Phi(1,5)=\frac{1}{2}+\frac{1}{2}\cdot 0,8664=0,9332;$$

$$P(X\ge 15,4)=1-F(15,4)=1-(\frac{1}{2}+\frac{1}{2}\Phi(\frac{15,4-15}{0,2}))=\frac{1}{2}-\frac{1}{2}\Phi(2)=\frac{1}{2}-\frac{1}{2}\cdot 0,9545=0,0228$$

$$P(14,9\le x\le 15,3)=\frac{1}{2}\left[\Phi(\frac{15,3-15}{0,2})-\Phi(\frac{14,9-15}{0,2})\right]=\frac{1}{2}\left[\Phi(1,5)+\Phi(0,5)\right]=$$

$$=\frac{1}{2}\cdot\left(0,3664+0,3829\right)=0,6246.$$

По “правилу трех сигм” $P(|X-15|\le 0,6)=0,9973$ и следовательно, $15-0,6\le X\le 15+0,6.$ Окончательно $14,4\le x\le 15,6$.

1.2.9. Биномиальное распределение

Биномиальным является распределение вероятностей появления m числа событий в n независимых испытаниях, в каждом из которых вероятность появления события постоянна и равна p. Вероятность возможного числа появлений события вычисляется по формуле Бернулли:

$$p(x = k) = C_n^k p^k q^{n-k}, \quad k = 0,1,...,n,$$

где $q = 1 - p$.

Постоянные n и p, входящие в это выражение, параметры биномиального закона. Биномиальным распределением описывается распределение вероятностей дискретной случайной величины.

Основные числовые характеристики биномиального распределения:

математическое ожидание равно $M(x) = n \cdot p$;

дисперсия равна $D(x) = n \cdot p \cdot q$;

коэффициент асимметрии $A = \dfrac{q - p}{\sqrt{npq}}$;

коэффициент эксцесса равен $E = \dfrac{1 - 6pq}{npq}$.

При неограниченном возрастании числа испытаний коэффициенты асимметрии и эксцесса стремятся к нулю, следовательно, можно предположить, что биномиальное распределение сходится к нормальному с возрастанием числа испытаний.

Пример 63. Производятся независимые испытания с одинаковой вероятностью появления события A в каждом испытании. Найти вероятность появления события A в одном испытании, если дисперсия числа появлений в трёх испытаниях равна 0,63.

Решение. Для биномиального распределения дисперсия равна $D(x) = n \cdot p \cdot q$. Подставим имеющиеся значения, получим $0{,}63 = 3p(1 - p)$.

Отсюда $p^2 - p + 0{,}21 = 0$. Решая квадратное уравнение, получим $p_1 = 0{,}7, \quad p_2 = 0{,}3$.

Пример 64. В городе 4 коммерческих банка. У каждого риск банкротства в течение года составляет 20%. Составьте ряд распределения числа банков, которые могут обанкротиться в течение следующего года.

Решение. Пусть *X* — дискретная случайная величина, равная числу банков, которые могут обанкротиться в течение следующего года. Она может принимать значения 0, 1, 2, 3 и 4. *X* распределена по биномиальному закону с параметрами и *n*= 4, *p* — 20% = 0,2, поэтому найдем соответствующие вероятности по формуле Бернулли:

$$P(x=k) = p_n(k) = C_n^k \cdot p^k \cdot (1-p)^{n-k} = \frac{n!}{k!(n-k)!} \cdot p^k \cdot (1-p)^{n-k}.$$

Получаем:

$$P(x=0) = p_4(0) = C_4^0 \cdot 0,2^0 \cdot 0,8^4 = 0,4096$$
$$P(x=1) = p_4(1) = C_4^1 \cdot 0,2^1 \cdot 0,8^3 = 0,4096$$
$$P(x=2) = p_4(2) = C_4^2 \cdot 0,2^2 \cdot 0,8^2 = 0,1536$$
$$P(x=3) = p_4(3) = C_4^3 \cdot 0,2^3 \cdot 0,8^1 = 0,0256$$
$$P(x=4) = p_4(4) = C_4^4 \cdot 0,2^4 \cdot 0,8^0 = 0,0016$$

Таким образом закон распределения случайной величины X имеет вид:

x_i	0	1	2	3	4
p_i	0,4096	0,4096	0,1536	0,0256	0,0016

Расчеты произведены правильно, так как сумма $\sum p_i = 1$.

Пример 65. Контрольная работа состоит из трех вопросов. На каждый вопрос приведено четыре варианта ответа, один из которых правильный. Составить закон распределения числа правильных ответов при простом угадывании. Найти $M(X), D(X)$.

Решение. Пусть *X* - дискретная случайная величина, равная числу правильных ответов. Она распределена по биномиальному закону с параметрами $n = 3$ (число вопросов), $p = \frac{1}{4}$ (вероятность угадать ответ), поэтому вероятности вычисляются но формуле Бернулли:

$$P(x=k)=p_n(k)=C_n^k\cdot p^k\cdot(1-p)^{n-k}=C_3^k\cdot(3/4)^k\cdot(1/4)^{3-k}.$$

Получаем:

$$P(X=3)=C_3^3\cdot(3/4)^3\cdot(1/4)^0=\frac{1}{64},$$
$$P(X=2)=C_3^2\cdot(3/4)^2\cdot(1/4)^1=\frac{9}{64},$$
$$P(X=1)=C_3^1\cdot(3/4)^1\cdot(1/4)^2=\frac{27}{64},$$
$$P(X=0)=C_3^0\cdot(3/4)^0\cdot(1/4)^3=\frac{27}{64}.$$

Получили закон распределения случайной величины X имеет вид:

x_i	0	1	2	3
p_i	27/64	27/64	9/64	1/64

Найдем числовые характеристики по формулам для биномиального распределения:

$$M(X)=np=3\cdot\frac{1}{4}=0,75,$$
$$D(X)=np(1-p)=3\cdot\frac{1}{4}\cdot\frac{3}{4}=0,5625.$$

Пример 66. На контроль качества медицинских препаратов поступила партия из $n=8$ штук. Вероятность того, что препарат окажется некачественным, равна 0,35.

А) Найти вероятности $p_n(k)$ того, что число некачественных препаратов k в партии составляет 0,1,…,8.

В) найти паи вероятнейшее число некачественных препаратов.

Решение. Дискретная случайная величина X — (Число некачественных препаратов в партии) распределена по биномиальному закону с параметрами $p=0,35, n=8,$ поэтому вероятности $P\left(X=k\right)$ найдем по формуле Бернулли:

$$P(x=k)=p_n(k)=C_n^k\cdot p^k\cdot(1-p)^{n-k}.$$

Получаем:

$$P(x=0)=p_8(0)=C_8^0\cdot 0,35^0\cdot(1-0,35)^8\approx 0,0319,$$
$$P(x=1)=p_8(1)=C_8^1\cdot 0,35^1\cdot(1-0,35)^7\approx 0,1373,...$$

Продолжая вычисления, получим ряд распределения:

k	$p_8(k)$
0	*0,0319*
1	*0,1373*
2	*0,2587*
3	*0,2785*
4	*0,1875*
5	*0,0808*
6	*0,0217*
7	*0,0033*
8	*0,0002*

Найдем наивероятнейшее число некачественных препаратов из таблицы. Наибольшая вероятность 0,2786 соответствует число препаратов, равному $k_0=3$.

Пример 67. Наблюдение за районом осуществляется тремя радиолокационными станциями (РЛС). В район наблюдений нормал объект, который обнаруживается любой радиолокационной станцией с вероятностью 0,2. Построить ряд распределения, найти функцию распределения, математическое ожидание и среднее квадратичное отклонение числа РЛС, обнаруживших объект.

Найти вероятность того, что их будет не менее двух.

Решение. Пусть *X*=(Число РЛС, обнаруживших объект). Она распределена но биномиальному закону распределения с вероятностью p = 0,2 (вероятность, что РЛС обнаружила объект), *n=3* (число РЛС) и q =1-*p* = 0,8. *X* может принимать значения 0,1,2 и 3.

Найдем соответствующие вероятности по формуле Бернулли:

$$P(x=k)=p_n(k)=C_n^k\cdot p^k\cdot(1-p)^{n-k}.$$

Получаем:

$$P(x=0)=C_3^0\cdot 0,2^0\cdot 0,8^3=0,512;$$
$$P(x=1)=C_3^1\cdot 0,2^1\cdot 0,8^2=0,384;$$
$$P(x=2)=C_3^2\cdot 0,2^2\cdot 0,8^1=0,096;$$
$$P(x=3)=C_3^3\cdot 0,2^3\cdot 0,8^0=0,008.$$

Таким образом, ряд распределения случайной величины *X* имеет вид:

x_i	0	1	2	3
p_i	0,512	0,384	0,096	0,008

Найдем функцию распределения $F(x)=P(X<x),$ то есть

при $x\le 0, F(x)=0,$

при $0<x\le 1,\ F(x)=0,512,$

при $1<x\le 2,\ F(x)=0,896,$

при $2<x\le 3,\ F(x)=0,992,$

при $3<x\le 4,\ F(x)=0,1.$

1.2.10. Распределение Пуассона

Распределение Пуассона описывает число событий m, происходящих за одинаковые промежутки времени при условии, что события происходят независимо друг от друга с постоянной средней интенсивностью (λ). При этом число испытаний n велико, а вероятность появления события в каждом испытании p мала. Поэтому распределение Пуассона называют законом редких явлений или простейшим потоком. Параметром распределения Пуассона является величина $\lambda=np$, характеризующая интенсивность появления событий в n испытаниях. Вероятность возможного числа появления события вычисляется по формуле распределения Пуассона

$$p_n(x=m)=\frac{\lambda^m}{m!}e^{-\lambda}.$$

Пуассоновским распределением хорошо описываются: число требований на выплату страховых сумм за год, число вызовов, поступивших на телефонную станцию за определённое время, число отказов элементов при испытании на надёжность, число бракованных изделий и так далее.

Основные числовые характеристики для распределения Пуассона.

Математическое ожидание совпадает с дисперсией и равно a. То есть $M(x) = D(x) = a$. Это является отличительной особенностью данного распределения.

Коэффициент асимметрии $A = \frac{1}{\sqrt{a}}$.

Коэффициент эксцесса $E = \frac{1}{a}$.

Пример 68. Среднее число выплат страховых сумм в день равно двум. Найти вероятность того, что за пять дней придётся выплатить: 1) шесть страховых сумм; 2) менее шести сумм; 3) не менее шести.

Решение. Среднее число выплат за пять дней $5 \cdot 2 = 10$.

Вероятность того, что придётся выплачивать шесть сумм, равна $p(k=6) = \frac{10^6}{6!}e^{-10} \approx 0{,}063055.$

Обозначим за событие A – выплачено менее шести сумм. Тогда по теореме сложения несовместных событий получим:

$$p(A) = p(k=0) + p(k=1) + p(k=2) + p(k=3) + p(k=4) + p(k=5) =$$

$$= \frac{10^0}{0!}e^{-10} + \frac{10^1}{1!}e^{-10} + \frac{10^2}{2!}e^{-10} + \frac{10^3}{3!}e^{-10} +$$

$$+ \frac{10^4}{4!}e^{-10} + \frac{10^5}{5!}e^{-10} \approx 0{,}067.$$

Событие $\overline{A}$ – выплачено не менее шести сумм является противоположным для события A, следовательно:

$$p(\overline{A}) = 1 - p(A) = 1 - 0{,}067 = 0{,}933.$$

Пример 69. Случайная величина X подчинена закону Пуассона с математическим ожиданием a = 3. Найти вероятность того, что данная случайная величина X примет значение, меньшее, чем ее математическое ожидание.

Отличие состоит в том, что здесь речь идёт именно о распределении Пуассона.

Решение. Случайная величина X принимает значения $x_i = \{0, 1, 2, ..., n, ..\}$ с вероятностями:

$$P_{x_i} = \frac{a^{x_i}}{x_i!} e^{-a}.$$

По условию, $M(x) = a = 3$, и тут всё просто: событие $X < 3$ состоит в трёх несовместных исходах:

$P(X < 3) = P_0 + P_1 + P_2 = \frac{3^0}{0!}e^{-3} + \frac{3^1}{1!}e^{-3} + \frac{3^2}{2!}e^{-3} = 0,4232$ того, что случайная величина x примет значение, меньшее, чем ее математическое ожидание.

Пример 70. В новом микрорайоне поставлено 10000 кодовых замков на входных дверях домов. Вероятность выхода из строя одного замка в течение месяца равна 0,0002. Найти вероятность того, что за месяц откажет ровно 1 замок.

Решение. В данном случае количество «испытаний» $n = 10000$ велико, а вероятность «успеха» в каждом из них - мала: p = 0,0002, поэтому используем формулу Пуассона:

$$P_m \approx \frac{\lambda^m}{m!} e^{-\lambda}.$$

Вычислим: $\lambda = np = 10000 \cdot 0,0002 = 2$ -по существу, это средне ожидаемое количество вышедших из строя замков.

Таким образом:

$P_1 \approx \frac{2^1}{1!} e^{-2} = 2e^{-2} \approx 0,2707$ - вероятность того, что за месяц из строя выйдет ровно m = 1 один замок (из 10 тысяч).

Пример 71. Завод отправил в торговую сеть 500 изделий. Вероятность повреждения изделия в пути равна 0,003. Найти вероятность того, что при транспортировке будет повреждено: а) ни одного изделия, б) ровно три изделия, в) более трех изделий.

Решение. Используем формулу Пуассона:

$$P_m \approx \frac{\lambda^m}{m!}e^{-\lambda}.$$

В данном случае $\lambda = np = 500 \cdot 0,003 = 1,5 -$ средне ожидаемое количество повреждённых изделий

а) *m=0*

$P_0 \approx \frac{1,5^0}{0!}e^{-1,5} \approx 0,2231$- вероятность того, что все изделия дойдут в целости и сохранности.

б) $m = 3$

$P_0 \approx \frac{1,5^3}{3!}e^{-1,5} \approx 0,1255$- вероятность того, что в пути будут повреждены ровно 3 изделия из 500.

$$P(m \le 3) \approx P_0 + P_1 + P_2 + P_3 = \frac{1,5^0}{0!}e^{-1,5} + \frac{1,5^1}{1!}e^{-1,5} + \frac{1,5^2}{2!}e^{-1,5} + \frac{1,5^3}{3!}e^{-1,5} \approx 0,9344$$

По теореме сложения вероятностей противоположных событий:

$P(m > 3) = 1 - P(m \le 3) \approx 1 - 0,9344 = 0,0665 -$ вероятность того, что при доставке будет повреждено более 3 изделий.

1.2.11. Показательное распределение

Непрерывную случайную величину, плотность вероятности которой определяется выражением

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

называют величиной, имеющей показательное или экспоненциальное распределение.

Это распределение часто наблюдается при изучении сроков службы различных устройств, времени безотказной работы отдельных элементов, частей системы и системы в целом, при рассмотрении случайных промежутков времени между появлениями двух последовательных редких событий.

Плотность показательного распределения определяется параметром λ, который называют интенсивностью отказов. Этот термин связан с конкретной областью приложения – теорией надёжности.

Выражение интегральной функции показательного распределения можно найти, используя свойства дифференциальной функции:

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

Основные числовые характеристики показательного распределения:

- математическое ожидание $M(x)=\dfrac{1}{\lambda}$;

- дисперсия $D(x)=\dfrac{1}{\lambda^2}$;

- среднее квадратическое отклонение $\sigma=\dfrac{1}{\lambda}$.

Таким образом, для этого распределения характерно, что среднее квадратическое отклонение численно равно математическому ожиданию. При любом значении параметра λ коэффициенты асимметрии и эксцесса – постоянные величины $A=2, \quad E=9$.

Пример 72. Среднее время работы телевизора до первого отказа равно 500 часов. Найти вероятность того, что наудачу взятый телевизор проработает без поломок более 1000 часов.

Решение. Так как среднее время работы до первого отказа равно 500, то $\dfrac{1}{\lambda}=500$, отсюда получим $\lambda=\dfrac{1}{500}$. Искомую вероятность найдём по формуле

$$p(t>1000)=\int_{1000}^{\infty}\frac{1}{500}e^{-\frac{t}{500}}dt=-e^{-\frac{t}{500}}\Big|_{1000}^{\infty}=e^{-2}\approx 0{,}135.$$

1.2.12. Закон больших чисел

Теорема 6. Для любой положительной случайной величины X вероятность того, что она примет значение, не превосходящее некоторого положительного числа τ, больше разности между единицей и отношением математического ожидания этой случайной величины к данному числу τ:

$$p(X \leq \tau) > 1 - \frac{M(x)}{\tau}.$$

Неравенство $p(X \leq \tau) > 1 - \dfrac{M(x)}{\tau}$ называется неравенством Маркова.

Пример 73. Среднее число студентов в группе, получивших на экзамене неудовлетворительные оценки, равно пяти. Оценить вероятность того, что в случайно взятой группе будет более восьми двоек.

Решение. Используем неравенство: $p(X > \tau) \leq \dfrac{M(x)}{\tau}$. По условию задачи имеем $M(x) = 5;\ \tau = 8$. Тогда искомая вероятность равна $p(x > 8) \leq \dfrac{5}{8} = 0{,}625$.

1.2.13. Неравенство Чебышева

Теорема 7. Если случайная величина X имеет конечные математическое ожидание и дисперсию, то для любого положительного числа τ справедливо неравенство

$$p(|x - M(x) \leq \tau|) > 1 - \frac{D(x)}{\tau^2},$$

то есть вероятность того, что отклонение случайной величины X от своего математического ожидания по абсолютной величине не превзойдёт τ, больше разности между единицей и отношением дисперсии этой случайной величины к квадрату τ.

Неравенство $p\left(\left|x - M(x) \le \tau\right|\right) > 1 - \frac{D(x)}{\tau^2}$ называется неравенством Чебышева. Неравенство Чебышева, как и неравенство Маркова, справедливо для любого распределения случайной величины X, однако неравенство Чебышева применимо как к положительным, так и к отрицательным случайным величинам.

Положим $\tau = t\sigma$, запишем неравенство в виде $p\left(\left|x - M(x) \le t\sigma\right|\right) > 1 - \frac{1}{\tau^2}$ или $p\left(\left|x - M(x)\right| > t\sigma\right) \le \frac{1}{t^2}$. Пусть $t = 3$, вычислим верхние границы вероятностей того, что отклонение случайной величины превысит 3σ, получим $p\left(\left|x - M(x)\right| > 3\sigma\right) \le \frac{1}{9}$.

Пример 74. Длина изготавливаемых деталей является случайной величиной, среднее значение которой 50 мм. Среднеквадратичное отклонение этой величины равно 0,2 мм. Оценить вероятность того, что отклонение длины изготовленной детали от ее среднего значения по абсолютной величине не превзойдет 0,4 мм.

Решение. Для оценки вероятности используем неравенство Чебышева

$$P(\left|X - M(X)\right| < \varepsilon) \ge 1 - \frac{D(X)}{\varepsilon^2},$$

$$P(\left|X - 50\right| < 0,4) \ge 1 - \frac{0,2^2}{0,4^2} = 1 - 0,25 = 0,75.$$

Ответ: $P(\left|X - 50\right| < 0,4) \ge 0,75$.

Пример 75. Среднесуточное потребление электроэнергии в населенном пункте равно 20000 кВт/ч, а среднеквадратичное отклонение — 200 кВт/ч. Какого потребления электроэнергии в этом населенном пункте можно ожидать в ближайшие сутки с вероятностью, не меньшей 0,96?

Решение. Воспользуемся неравенством Чебышева

$$P(\left|X - M(X)\right| < \varepsilon) \ge 1 - \frac{D(X)}{\varepsilon^2}.$$

Подставим в правую часть неравенства в место $D(X)$ величину

$200\wedge 2 = 40000$, сделаем ее большей или равной 0,96:

$$1-\frac{40000}{\varepsilon^2}\geq 0,96 \Leftrightarrow \frac{40000}{\varepsilon^2}\leq 0,04 \Leftrightarrow \varepsilon^2 \geq \frac{40000}{0,04},\ \varepsilon \geq 1000.$$

Следовательно, в этом населенном пункте можно ожидать с вероятностью не меньшей 0,96 потребление электроэнергии

$$20000 \pm 1000, т.е. X \in [19000; 21000].$$

Ответ: от 19000 до 21000.

1.2.14. Теорема Чебышева

Теорема 8. При достаточно большом числе попарно независимых случайных величин $X_1, X_2, ..., X_n$ с математическими ожиданиями $M(x_1), M(x_2), ..., M(x_n)$ и дисперсиями $D(x_1), D(x_2), ..., D(x_n)$ с вероятностью, близкой к единице, можно утверждать, что разность между средней арифметической наблюдавшихся значений случайных величин $X_1, X_2, ..., X_n$ и средней арифметической их математических ожиданий по абсолютной величине окажется меньше сколь угодно малого числа $\tau > 0$ при условии, что дисперсия всех этих величин не превосходит одного и того же постоянного числа B, то есть

$$p\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}M(x_i)\right| \leq \tau\right) > 1-\varepsilon,$$

где ε – положительное число, близкое к нулю.

Теорема Чебышева показывает, что средняя арифметическая попарно независимых случайных величин обладает свойством устойчивости и при определённых условиях мало отличается от средней арифметической математических ожиданий этих величин.

Пример 76. За значение некоторой величины принимают среднеарифметическое достаточно большого числа ее измерений. Предполагая, что среднеквадратичное отклонение возможных результатов каждого измерения не превосходит 5 мм, оценить вероятность того, что при 1000 измерений неизвестной величины отклонение принятого значения от истинного по абсолютной величине не превзойдет 0,5 мм.

Решение. Воспользуемся неравенством

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}X_i\right|<\varepsilon\right)\geq 1-\frac{C}{n\varepsilon^2}.$$

По условию $n=1000, \varepsilon=0,5, C=5^2=25$. Итак, искомая вероятность

$$P\left(\left|\frac{1}{1000}\sum_{i=1}^{1000}X_i-\frac{1}{1000}\sum_{i=1}^{1000}M(X_i)\right|<0,5\right)\geq 1-\frac{25}{1000\cdot 0,25}=0,9.$$

Ответ $P\geq 0,9$.

Частными случаями теоремы Чебышева являются теоремы Бернулли и Пуассона.

Теорема Бернулли. При неограниченном увеличении числа независимых опытов частость появления $\frac{m}{n}$ Некоторого события A сходится по вероятности к его вероятности $P=P(A)$:

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

где $\varepsilon-$ сколь угодно малое положительное число.

При доказательстве теоремы Бернулли получаем такую оценку

$$\lim_{x\to\infty}P(\left|\frac{m}{n}-p\right|<\varepsilon)\geq 1-\frac{pq}{n\varepsilon^2},$$

которая применяется на практике.

Теорема Пуассона. Если производится n независимых опытов и вероятность появления события A сходиться по вероятности средне арифметическому вероятностей p_k:

$$\lim_{x\to\infty}P(\left|\frac{m}{n}-\frac{1}{n}\sum_{k=1}^{n}p_k\right|<\varepsilon)=1,$$

где $\varepsilon-$ сколь угодно малое положительное число. При доказательстве этой теоремы используется неравенство

$$P(\left|\frac{m}{n}-\frac{1}{n}\sum_{k=1}^{n}p_k\right|<\varepsilon)\geq 1-\frac{1}{4n\varepsilon^2},$$

имеющее практическое применение.

Пример 77. При контрольной проверке изготавливаемых приборов было установлено, что в среднем 15 шт. из 100 оказывается с теми или иными дефектами. Оценить вероятность того, что доля приборов с дефектами среди 400 изготовленных будет по абсолютной величине отличаться от математического ожидания этой доли не более чем на 0,05.

Решение. Воспользуемся неравенством

$$\lim_{x\to\infty} P(\left|\frac{m}{n}-p\right|<\varepsilon)\geq 1-\frac{pq}{n\varepsilon^2}.$$

По условию $n=400, \varepsilon=0,05.$ В качестве P возьмем величину, полученную при проверке для доли брака $p=\frac{15}{100}=0,15$.

Итак,

$$P(\left|\frac{m}{n}-p\right|<\varepsilon)\geq 1-\frac{0,15\cdot 0,85}{400\cdot 0,05^2}=0,8725.$$

Ответ $P\geq 0,8725.$

Пример 78. Вероятность того, что изделие является качественным, равна 0,9. Сколько следует проверить изделий, чтобы с вероятностью не меньшей 0,95 можно было утверждать, что абсолютная величина отклонения доли качественных изделий от 0,9 не превысит 0,01?

Решение. Воспользуемся неравенством

$$\lim_{x\to\infty} P(\left|\frac{m}{n}-p\right|<\varepsilon)\geq 1-\frac{pq}{n\varepsilon^2}.$$

По условию $p=0,9, q=1-0,9=0,1, \varepsilon=0,01.$ Подставим в правую часть вышеприведенного неравенства эти значения

$$1-\frac{0,9\cdot 0,1}{n\cdot 0,0001}\geq 0,95 \Leftrightarrow \frac{900}{n}\leq 0,05 \Leftrightarrow n\geq 18000.$$

Ответ: $n\geq 18000.$

Задачи для самостоятельного решение

1.1. Сколькими способами можно рассадить 5 человек за столом?

1.2. Сколько четырёхзначных чисел можно составить из четырёх карточек с цифрами 0, 5, 7, 9?

1.3. В шахматном турнире участвует k человек и каждый с каждым играет по одной партии. Сколько всего партий сыграно в турнире?

1.4. В студенческой группе 23 человека. Сколькими способами можно выбрать старосту и его заместителя?

1.5. Студенческая группа состоит из 23 человек, среди которых 10 юношей и 13 девушек. Сколькими способами можно выбрать двух человек одного пола?

1.6. У Васи дома живут 4 кота.

а) сколькими способами можно рассадить котов по углам комнаты?
б) сколькими способами можно отпустить гулять котов?
в) сколькими способами Вася может взять на руки двух котов (одного на левую, другого – на правую)?

1.7. В лифт 12-этажного дома сели 3 пассажира. Каждый независимо от других с одинаковой вероятностью может выйти на любом (начиная со 2-го) этаже. Сколькими способами:

1) пассажиры могут выйти на одном и том же этаже *(порядок выхода не имеет значения)*;
2) два человека могут выйти на одном этаже, а третий – на другом;
3) люди могут выйти на разных этажах;
4) пассажиры могут выйти из лифта?

1.8. Сколько различных буквосочетаний можно получить перестановкой карточек со следующими буквами: К, О, Л, О, К, О, Л, Ь, Ч, И, К?

1.9. В автомашине 7 мест. Сколькими способами 7 человек могут усесться в эту машину, если водительские права имеют только трое из них?

1.10. Из двадцати человек, которые должны сдавать экзамены, 10 должны явиться в 10 часам утра, остальные - к 11 часам. Если 7 студентов определенно хотят быть в первой группе, 5 - во второй, а две подружки не возражают быть в любой из групп, но только обязательно вместе, то сколькими способами староста может распределить студентов по группам?

1.11. В кондитерской имеется 7 видов пирожных. Сколько есть способов заказать 4 пирожных? Сколько среди них есть способов заказать пирожное одного вида? Разных видов? По 2 пирожных разных видов?

1.12. Комитет состоит из 12 членов. Минимальный кворум на заседаниях этого комитета должен насчитывать 8 членов.

а) Сколькими способами может достигаться минимальный форум?

б) Сколькими способами достигается какой-нибудь кворум?

1.13. На восьми одинаковых карточках написаны соответственно числа 2, 4, 6, 7, 8, 11, 12, 13. Наугад берутся две карточки. Определить вероятность того, что образованная из двух полученных чисел дробь несократима.

1.14. Бросаются две игральные кости. Пусть A - событие, состоящее в том, что сумма очков нечетная, B - событие, заключающееся в том, что хотя бы на одной из костей выпала единица. Составить пространство элементарных событий, связанное с данным опытом.

1.15. Потребитель может увидеть рекламу определенного продукта по телевидению, на рекламном стенде и прочитать в газете. Составить пространство элементарных событий для потребителя в этом опыте.

1.16. Торговый агент последовательно контактирует с тремя потенциальными покупателями. Под исходом опыта будем понимать последовательность (X_1, X_2, X_3), где каждый из X_i обозначает продажу или нет $(\overline{X_i})$ товара покупателю. Построить пространство Ω элементарных событий.

1.17. Из таблицы чисел взято число. Событие A - число делится на 5, событие B - число оканчивается нулем. а) Что означают события $A-B$ и $A\overline{B}$? б) Совместны ли события A и $\overline{A+B}$?

1.18. Из множества супружеских пар наугад выбирается одна пара Событие A «Мужу больше 30 лет», событие B «Муж старше жены», событие C: «Жене больше 30 лет».

1. Выяснить смысл событий ABC, $A-AB$, $A\bar{B}C$.

2. Проверить, что $A\bar{C}\subset B$.

1.19. Рабочий обслуживает три автоматических станка Событие A — первый станок потребует внимания рабочего в течение часа, B — второй станок потребует внимания рабочего в течение часа, C — третий станок потребует внимания рабочего в течение часа. Что означают события

а) ABC, б) $A+B+C$. в) $A\bar{B}\bar{C}+\bar{A}B\bar{C}+\bar{A}\bar{B}C$ г) $AB\bar{C}+A\bar{B}C+\bar{A}BC$

е) $A+B+C-ABC$?

1.20. Производится испытание трех приборов на надежность Пусть событие $A_k - k$-й прибор выдержал испытание $\{k=1,2,3\}$. Представить в виде суммы и произведения события A_k и $\overline{A_k}$ следующие события а) хотя бы один прибор выдержал испытание, б) не менее двух приборов выдержали испытание, в) только один прибор выдержал испытание, г) только два прибора выдержали испытание.

1.21. Страховой агент предлагает услугу по страхованию жизни трем потенциальным клиентам. Пусть события A, B и C означают соответственно, что первый, второй и третий клиент согласился застраховать свою жизнь.

1) Составить события:

а) все клиенты согласились на страховку,

б) хотя бы один клиент согласился на страховку,

в) только один клиент согласился на страховку,

в) только первый клиент согласился на страховку

2) Назвать события

а) $AB\bar{C}+A\bar{B}C+\bar{A}BC$ б) $A\overline{BC}$ с) $\bar{A}+\bar{B}+\bar{C}$ г) $\overline{ABC}$

1.22. Два игрока поочередно бросают монету Выигрывает тот игрок, у которого раньше выпадет герб. Пусть A_l означает событие, что в l-ой партии у первого игрока выпал герб. B_i - у второго игрока в l-ой партии выпал герб, A - выигрыш первого игрока. B - выигрыш второго игрока Записать выражение для A и B через A_i, B_i, $\overline{A_i}$, $\overline{B_i}$, $i = 1, 2, \ldots$

1.23. Если событие A - выигрыш по билету одной лотереи, B - выигрыш по билету другой лотереи, то что означают события $C = A\overline{B} + \overline{A}B,\ D = A\overline{B} + \overline{A}B + AB$?

1.24. На пяти карточках написаны цифры от 1 до 5 Опыт состоит в случайном выборе трех карточек и раскладывании их в порядке поступления в ряд слева направо. Найти вероятность следующих событий $A =$ (появится число 123). $B =$ (появится число, не содержащее цифры 2). $C =$ (появится число, состоящее из последовательных цифр).

1.25. Десять человек входят в комнату, где имеется всего 7 стульев, и рассаживаются случайным образом, но так, что все стулья оказываются занятыми. Какова вероятность того, что а) два определенных лица окажутся без места? б) 4 определенных лица будут сидеть?

1.26. Фирмы $A1, A2, A3, A4, A5$ предлагают свои условия по выполнению 3 различных контрактов $C1, C2, C3$. Любая фирма может получить только один контракт. Если предположить равно возможность заключения контрактов, чему равна вероятность того, что фирма $A3$ получит контракт? Чему равна вероятность того, что фирмы $A1$ и $A2$ получат контракт?

1.27. 8 вариантов контрольной работы, написанные каждый на отдельной карточке, перемешиваются и распределяются случайным образом среди шести студентов, сидящих в одном ряду, причем каждый получает по одному варианту Найти вероятность следующих событий $A =$ «варианты с номерами 1 и 2 останутся неиспользованными». $B =$ «варианты 1 и 2 достанутся рядом сидящим студентам». $C =$ «будут распределены последовательные номера вариантов».

1.28. A и B и еще 8 человек стоят в очереди. Определить вероятность того, что . A и B отделены друг друга тремя лицами.

1.29. Пять мужчин и пять женщин случайным образом рассаживаются в ряд на десять мест. Найти вероятности следующих событий A = «никакие два мужчины не будут сидеть рядом». B = «все мужчины будут сидеть рядом». C = «мужчины и женщины будут чередоваться».

1.30. В лифт семиэтажного дома на первом этаже вошли 3 человека Каждый из них с одинаковой вероятностью выходит на любом из этажей, начиная со второго Найти вероятности следующих событий A = «все пассажиры выйдут на четвертом этаже». B = «все пассажиры выйдут одновременно (на одном и том же этаже)»; С = «все пассажиры выйдут на разных этажах».

1.31. 9 пассажиров наудачу рассаживаются в трех вагонах. Найти вероятность того, что а) в каждый вагон сядут по три пассажира. 6) в один вагон сядут 4. в другой - 3. в третий - 2 пассажира.

1.32. На плоскость с нанесенной сеткой квадратов со стороной a наудачу брошена монета радиуса $r < a/2$. Найти вероятность того, что монета не пересечет ни одной из сторон квадрата. Предполагается, что вероятность попадания точки в плоскую фигуру пропорциональна площади фигуры и не зависит от ее расположения.

1.33. На плоскость, разграфленную параллельными прямыми, отстоящими друг от друга на расстоянии $6\,см$, наудачу брошен круг радиуса $1\,см$. Найти вероятность того, что круг не пересечет ни одной из прямых. Предполагается, что вероятность попадания точки на отрезок пропорциональна длине отрезка не зависит от его расположения.

1.34. Из чисел 1, 2, $\ldots$, 49 наугад выбираются и фиксируются 6 чисел, считающиеся выигрышными. Некто, желающий выиграть, наугад называет свои 6 чисел из 49. Какова вероятность, что среди названных им чисел окажется не менее трех выигрышных?

1.35. Группа, состоящая из $3n$ юношей и 3 девушек, делится произвольным образом на три равные по количеству подгруппы. Какова вероятность, что все девушки окажутся в разных подгруппах?

1.36. n студентов произвольным образом расходятся по k аудиториям. Какова вероятность, что в первой аудитории окажется n_1 студентов, во второй - n_2 студентов в k-й аудитории - n_k студентов, $n_1 + ... + n_k = n$?

1.37. В купейный вагон (9 купе но 4 места) продано $№ + 4$ билетов. Найти вероятность того, что занятыми оказались ровно пять купе (через № обозначен номер студента но списку группы).

1.38. Из отрезка $[0,1)$ наугад выбирается число. Какова вероятность, что в десятичной записи этого числа вторая цифра после запятой будет двойкой?

1.39. В квадрат с вершинами (0. 0). (0, 1), (1, 0). (1, 1) наудачу брошена точка. Обозначим ее координаты через $X; Y$. Предполагается, что вероятность попадания в область, лежащую целиком внутри квадрата, зависит лишь от площади этой области и пропорциональна ей:

а) доказать, что для $0 < u < 1,\ 0 < v < 1$ выполнено

$$P\{X < u, Y < v\} = P\{X < u\}P\{Y < v\} = uv$$

б) найти для $0 < t < 1$ вероятности

$$1)\ P\{|X - Y| < t\}\};\ 2)\ P\{XY < t\};$$

$$3)\ P\{\max(X,Y) < t\};\ 4)\ P\{\min(X,Y) < t\};$$

в) найти $P\{X + Y < t\}\}$ для $0 < t < 2$.

1.40. На отрезок длины l произвольным образом брошены три точки. Пусть X, Y, Z — расстояния до этих точек от левого конца отрезка. Какова вероятность, что из отрезков с длинами X, Y и Z можно составить треугольник?

1.41. На отрезок единичной длины произвольным образом брошены две точки, которые делят отрезок на три части. Какова вероятность, что из этих частей можно составить треугольник?

1.42. На плоскости начерчены две концентрические окружности, радиусы которых 5 и 10 см соответственно. Найти вероятность того, что точка, брошенная наудачу в большой круг, попадет также и в кольцо, образованное построенными окружностями. Предполагается, что вероятность попадания точки в плоскую фигуру пропорциональна площади этой фигуры и не зависит от ее расположения.

1.43. На отрезке OA длины L числовой оси Ox наудачу поставлены две точки $B(x)$ и $C(y)$. Найти вероятность того, чт1о длина отрезка BC меньше расстояния от точки O до ближайшей к ней точке. Предполагается, что вероятность попадания точки на отрезок пропорциональна длине отрезка и не зависит от его расположения на числовой оси.

1.44. На отрезке OA длины L числовой оси Ox наудачу поставлены две точки: $B(x)$ и $C(y)$, причем $y \geq x$. Найти вероятность того, что длина отрезка BC окажется меньше, чем $L/2$. Предполагается, что вероятность попадания точки на отрезок пропорциональна длине отрезка и не зависит от его расположения на числовой оси.

1.45. На отрезке OA длины L числовой оси Ox наудачу поставлены две точки $B(x)$ и $C(y)$. Найти вероятность того, что длина отрезка BC окажется меньше, чем $L/2$. Предполагается, что вероятность попадания точки на отрезок пропорциональна длине отрезка и не зависит от его расположения на числовой оси.

1.46. В течение года фирмы *A, B, C,* независимо друг от друга, могут обанкротиться с вероятностями 0,06; 0,09 и 0,05 соответственно. Найти вероятности того, что к концу года 1) все три фирмы будут функционировать, 2) все три фирмы обанкротятся, 3) только одна фирма обанкротится. 4) только две фирмы обанкротятся, 5) хотя бы одна фирма обанкротится.

1.47. Пусть вероятность того, что в секции магазина по продаже мужской обуви очередной будет продана пара обуви 44-го размера, равна 0,12, 45-го — 0,04, 46-го или большего — 0,01. Найти вероятность того, что очередной будет продана пара мужской обуви не менее 44-го размера.

1.48. Студент выучил к зачету 15 вопросов из 20. Ему по одному предлагают три вопроса. Найти вероятность того, что только на третий из них он не дает ответа.

1.49. Для рабочего из маршрутов трамвая № 1, 2, 4, 7 попутными являются маршруты № 1 и 4. Вычислить вероятность того, что к остановке первым подойдет трамвай маршрута попутного для него номера, если по линиям маршрутов № 1, 2,4,7 курсируют соответственно 12,4, 10, 14 поездов.

1.50. Два охотника стреляют в волка Для первого охотника вероятность попадания в цель 0,7, для второго — 0,8. Определить вероятность попадания в волка, если каждый охотник 1) делает по одному выстрелу, 2) делает по два выстрела?

1.51. Для сигнализации об аварии установлены два независимо работающих сигнализатора Вероятность того, что при аварии сигнализатор сработает, равна 0,95 для первого сигнализатора и 0,9 для второго. Найти вероятность того, что при аварии сработает а) только один сигнализатор, б) хотя бы один сигнализатор.

1.52. Для компании, занимающейся строительством терминалов для аэропортов, вероятность получить контракт в стране A равна 0,4, вероятность выиграть его в стране B равна 0,3. Вероятность того, что контракты будут заключены и в стране A, и в стране B, равна 0,12. Какова вероятность того, что компания получит контракт: а) хотя бы в одной стране; б) только в одной стране?

1.53. Обследовалась группа из 10000 человек в возрасте свыше 60 лет. Оказалось, что 4 000 из них постоянно курит. У 1 800 человек из курящих обнаружились серьезные изменения в легких. Среди некурящих серьезные изменения в легких имели 1 500 человек. Являются ли курение и наличие серьезных изменений в легких независимыми событиями? (Ответ дать, проверив выполнение равенства $P(AB) = P(A)P(B)$ где событие A - человек курит, событие B - человек имеет серьезные изменения в легких.

1.54. О двух акциях A и B известно, что они выпущены одной и той же отраслью Вероятность того, что акция Л поднимется завтра в цене, равна 0,2. Вероятность того, что обе акции A и B поднимутся завтра в цене, равна 0,12. Предположим, что вы знаете, что акция A поднимется в цене завтра Чему равна вероятность того, что и акция B завтра поднимется в цене?

1.55. В фирме 550 работников, 380 из них имеют высшее образование, а 412 — среднее специальное образование, у 357 высшее и среднее специальное образование. Чему равна вероятность того, что случайно выбранный работник имеет: а) хотя бы одно из этих образований, б) только одно из этих образований, в) работник имеет только среднее специальное образование.

1.56. Вероятности успешной сдачи сессии у студентов Иванова и Петрова равны соответственно 0,95 и 0.9. Найти вероятности следующих событий

а) оба студент успешно сдадут сессию,

б) Иванов сдаст сессию успешно, а Петров не сдаст,

в) только один из студентов сдаст сессию успешно

Предполагается, что Иванов и Петров независимо друг от друга готовятся к сессии Ответ: а) $P=0{,}855$; б) $P=0{,}095$; в) $P=0{,}14$

1.57. Покупатель может приобрести акции двух компаний A и $Б$. Надежность первой компании оценивается экспертами на уровне 90 %, а второй — 80 % Чему равна вероятность того, что а) обе компании в течение года не станут банкротами, б) только одна компания станет банкротом, в) наступит хотя бы одно банкротство?

1.58. В автопробеге участвуют 3 автомобиля. Первый может сойти с маршрута с вероятностью 0,15, второй - с вероятностью 0,05, а третий - с вероятностью 0,1. Определить вероятность того, что к финишу придут: а) только один автомобиль, б) два автомобиля, в) по крайней мере два автомобиля.

1.59. Радист посылает вызов корреспонденту до тех пор, пока тот его не услышит, но при этом может послать не более трех вызовов Вероятность того, что корреспондент примет первый вызов, равна 0.2. второй — 0,3, третий — 0,4 По условиям приема события, состоящие в том, что l-й по счету вызов ($l=1,2,3$) услышан, независимы. Найти вероятность того, что корреспондент вообще услышит радиста.1. Устройство содержит два независимо работающих элемента. Вероятности отказа элементов соответственно равны 0,05 и 0,08. Найти вероятности отказа устройства, если для этого достаточно, чтобы отказал хотя бы один элемент.

1.60. Для разрушения моста достаточно попадания одной авиационной бомбы. Найти вероятность того, что мост будет разрушен, если на него сбросить четыре бомбы, вероятности попадания которых соответственно равны: 0,3; 0,4; 0,6; 0,7.

1.61. Вероятность успешного выполнения упражнения для каждого из двух спортсменов равна 0,5. Спортсмены выполняют упражнение по очереди, причем каждый делает по две попытки. Выполнивший упражнение первым получает приз. Найти вероятность получения приза спортсменами.

1.62. Вероятность попадания в мишень каждым из двух стрелков равна 0,3. Стрелки стреляют по очереди, причем каждый должен сделать по два выстрела. Попавший в мишень первым получает приз. Найти вероятность того, что стрелки получат приз.

1.63. Вероятность хотя бы одного попадания стрелком в мишень при трех выстрелах равна 0,875. Найти вероятность попадания при одном выстреле.

1.64. Вероятность хотя бы одного попадания в цель при четырех выстрелах равна 0,9984. Найти вероятность попадания в цель при одном выстреле.

1.65. На двух автоматических линиях изготовляют одинаковые детали: на первой- 30%, на второй- 70%. Вероятность изготовления стандартной детали на первой линии является 0,9, на второй-0,5. Все изготовленные на этих линиях детали наступают на склад.

а) Найти вероятность того, что наугад выбранная со склад деталь стандартная.

б) Наугад выбранная деталь, изготовленная на одной из линий, оказалось стандартной, Найти вероятность того, сто она изготовлена на первой линии.

1.66. В первой урне 4 белых и 3 чёрных шара, а во второй- 3 белых и один чёрный шар. Из первой урны наугад вынули один шар и переложили его во вторую урну. Найти вероятность того, что после перекладывания наугад вынутый из второй урны шар будет белым.

1.67. В пирамиды пять винтовок, три из которых снабжены оптическом прицелом. Вероятность того, что стрелок поразить мишень при выстреле из винтовки с оптическом прицелом, равна 0,95; для винтовки без оптического прицела эта вероятность равна 0,7. Найти вероятность, того что мишень будет поражена, если стрелок произведет один выстрел из наудачу взятой винтовки.

1.68. В урну содержат n шаров, опущен белый шар, после чего наудачу извлечен один шар. Найти вероятность того, что извлеченный шар оказаться белым, если рано возможны все возможны предположения о первоначальном составе шаров (по цвету).

1.69. Два автомата производят одинаковые детали, которые поступают она общий конвейер. Производительность первого автомата вдвое больше производительности второго. Первый автомат производить в среднем 60% деталей отличного качества, а второй - 84%. Найти удачу взятая с конвейера деталь оказалась отличного качества. Найти вероятность того, что эта деталь произведена первым автоматом.

1.70. В первой урне содержится 10 шаров, из них 8 белых; во второй урне 20 шаров, из них 4 белых. Из каждой урны наудачу извлекли по одному шару, а затем из этих двух шаров наудачу взят один шар. Найти вероятность того, что взят белый шар.

1.71. Число грузовых автомашин, проезжающих по шоссе, на котором стоит бензоколонка, относится к числу легковых машин, проезжающих по тому же шоссе как 3:2. Вероятность того, что будет заправляться грузовая машина, равна 0,1; для легковой машины эта вероятность рана 0,2. К бензоколонке подъехала для заправки машина. Найти вероятность того, что это грузовая машина.

1.72. В урне 8 шаров – 5 белых и 3 чёрных. Наугад вынимают два шара. Найти вероятность того, что второй шар оказался чёрным, если :

а) первый шар возвращали в урну;

б) первый шар не возвращали в урну.

1.73. Из 30 студентов группы, которые пришли на экзамен, 8 - подготовлены отлично, 10 - хорошо, 8 - удовлетворительно, а остальные - неудовлетворительно. Программа экзамена включает 50 вопросов. Билет содержит 3 вопроса. Студент, подготовленный отлично, знает все вопросы; хорошо – 40 вопросов; удовлетворительно – 25 вопросов и не удовлетворительно – 10 вопросов.

а) Найти вероятность того, что наугад вызванный студент ответит на все 3 вопроса билета.

б) Студент ответить на все вопросы. Найти вероятность того, что студент подготовлен:

- хорошо;
- удовлетворительно;
- неудовлетворительно.

1.74. В центр статических исследований поступает информация их трёх пунктов: первого – 50 %, второго – 30%, третьего – 20 % всей информации. Вероятность допущения ошибки при обработке статических данных в первом пункте равняется 0,1, во втором – 0,05, в третьем-0,15. Какая вероятность того, что полученная центром в данный момент информация правильная?

1.75. В одном классе 5 отличников, в втором - 3 отличника, а в третьем классе отличников нет. Из наугад взятого класса выбрали ученика. Найти вероятность того, что он отличник, если в каждом классе по 30 учеников.

1.76. Два экономиста заполняют документа, которые складывают в общую папку. Вероятность, допустить ошибку для первого экономиста равняется 0,1, для второго - 0,2. Первый экономист заполнил 40 документов, второй – 60. Во время проверни наугад взятый из папки документ оказался с ошибкой. Найти вероятность того, что его заполнил первый экономист.

1.77. Рабочий обслуживает три станка. Вероятность изготовления бракованной детали на первом станке равняется 0,04, на втором – 0,01, на третьем –

0,07. Производительность работы первого станка втрое превышает производительность второго, а производительность третьего вдвое меньше производительности второго.

а) Найти вероятность того, что наугад выбранная деталь, изготовленная рабочим, бракованная.

б) Взятая наугад деталь оказалось бракованной. На каком станке, вероятнее всего, она изготовлена?

1.78. Вероятности того, что во время работы компьютера произойдет собой в арифметическом устройстве, в оперативной памяти или в устройстве ввода соотносятся как 2:1:3. Вероятность обнаружить сбой в этих устройствах равна соответственно 0,9;0,75;0,7. Найти вероятность обнаружения сбоя в работе компьютера.

1.79. В корзине 20 яблок зелёного и 15 красного цвета. Наугад берут три яблока из корзины и кладут вместо них три яблока красного цвета. Потом наугад вынимают одно яблоко. Какая вероятность того, что вынуто яблоко цвета?

1.80. Изделие проверяется на стандартность одним из двух товароведов. Вероятность того, что изделие попадает к первому товароведу, равна 0,55, а ко второму – 0,45. Вероятность того, что стандартное изделие будет признано стандартным первым товароведом, равна 0,9, а вторым – 0,98. Стандартное изделие при проверке было признано стандартным. Найти вероятность того, что это изделие проверил второй товаровед.

1.81. Батарея из трех орудий произвела залп, причем два снаряда попали в цель. Найти вероятность того, что первое орудие дало попадание, если вероятности попадания в цель первым, вторым и третьим орудиями соответственно равны $p_1 = 0.4,\ p_2 = 0.3,\ p_3 = 0.5.$

1.82. Два из трех независимо работающих элементов вычислительного устройства отказали. Найти вероятность того, что отказали первый и второй элементы, если вероятности отказа первого, второго и третьего элементов соответственно равны 0,2; 0,4 и 0,3.

1.83. Стрелок А поражает мишень с вероятностью 0,6, стрелое В – с вероятностью 0,5, стрелок С с вероятнстью 0,4. Стрелки дали залп по мишени. Какова вероятность, что ровно две пули попали в цель?

1.84. Двое играют в игре, поочередно бросая монету. Выигравшим считается тот, кто первым получить герб. Найти вероятность того, что игра закончится на k-м бросании. Какова вероятность выигрыша для игрока, начинающего игру?

1.85. 10 любители подледного лова рыбы независимо друг от друга произвольным образом размещаются на льду озера, имеющего форму круга радиуса 1 км. Какова вероятность того, что не менее 5 рыбаков располагается на расстоянии более 200 м от берега?

1.86. Найти вероятность того, что в nиспытаниях схемы Бернулли с вероятностью успеха p появятся $m+l$ успехов, причем l успехов появятся в последних l испытаниях.

1.87. В круг вписан квадрат. Найти вероятность того, что из 10 точек, брошенных наудачу в круг, четыре попадут в квадрат, три – в нижней сегмент, и по одной – в оставшиеся три сегмента.

1.88. Найти вероятность того, что k-й по порядку успех в серии последовательных испытаний Бернулли произойдет на l-м испытании.

1.89. На отрезок $[0, 10]$ наудачу брошено5 точек. Найти вероятность того, что две точки попадут в $[0, 2]$, одна – в $[2,3]$ и две – в $[3,10]$.

1.90. Найти вероятность того, что при четырех подбрасываниях игральной кости 5 очков появится: а) два раза; б) хотя бы один раз; в) не менее 3 раз.

1.91. Всхожесть семян некоторого растения составляет 80%. Найти вероятность того, что из пяти посеянных семян взойдут: а) пять семян; б) не менее четырех; в) не более одного.

1.92. Два равносильных противника играют в шахматы. Что вероятнее а) выиграть одну партию из двух или две парты из четырех? Б) выиграть не менее двух партий из четырех или не менее трех партий из пяти? Ничьи во внимание не принимаются.

1.93. Монету бросают пять раз. Найти вероятность того, что «Герб» выпадает: а) менее двух раз; б) не менее двух раз.

1.94. а) Найти вероятность того, что событие А появится не менее трех раз в четырех независимых испытаниях, если вероятность появления события А в одном испытании равна 0,4;

б) событие B появится в случае, если события А наступит не менее четырех раз. Найти вероятность наступления события B, если будет произведено пять независимых испытаний, в каждом из которых вероятность появления события А равна 0,8.

1.95. Вероятность рождения мальчика равна 0,51. Найти вероятность того, что среди 100 новорожденный» окажется 50 мальчиков.

1.96. Монета брошена $2N$ раз (N велико!). Найти вероятность того, что «герб» выпадет ровно N раз.

1.97. Монета брошена $2N$ раз. Найти вероятность того, что «герб» выпадет на $2m$ раз больше, чем надпись.

1.98. Вероятность появления события в каждом из 100 независимых испытаний постоянна и равна р = 0,8. Найти вероятность того, что событие появится: а) не менее 75 раз и не более 90 раз; б) не менее 75 раз; В) не более 74 раз.

1.99. Вероятность появления события в каждом из 2100 независимых испытаний равна 0,7. Найти вероятность того, что событие появится: а) не менее 1470 и не более 1500 раз; б) не менее 1470 раз; в) не более 1469 раз.

1.100. Вероятность появления события в каждом из 21 независимых испытаний равна 0,7. Найти вероятность того, что событие появится в большинстве испытаний.

1.101. Монета брошена $2N$ раз (N велико!). Найти вероятность того, что число выпадений «герба» будет заключено между числами

$$N-\sqrt{2N}/2 \ \ и \ \ N+\sqrt{2N}/2.$$

1.102. Вероятность появления события в каждом из независимых испытаний равна 0,8. Сколько нужно произвести испытаний, чтобы с ероятностью 0,9 можно было ожидать, что событие появится не менее 75 раз?

1.103. Обзорную лекцию должны прослушать 100 студентов. Вероятность присутствия на этой лекции для каждого студента равняется 0.7. Найти вероятность того, что на лекцию пройдёт больше половины студентов.

1.104. Вероятность своевременной реализации со склада одной нары обуви равняется 0.8. Найти вероятность того, что своевременно будет реализовано не менее 75 пар. если на склад завезено 100 нар обуви. Найти наиболее вероятное количество своевременно реализованных пар обуви.

1.105. Среднее число заказов такси, поступающих на диспетчерский пункт в одну минуту, равно трем. Найти вероятность того, что за 2 мин поступит: а) четыре вызова; б) менее четырех вызовов; в) не менее четырех вызовов.

1.106. Среднее число вызовов, поступающих на АТС в одну минуту, равно двум. Найти вероятность того, что за 4 мин поступит: а) три вызова; б) менее трех вызовов; в) не менее трех вызовов. Поток вызовов предполагается простейшим.

1.107. Из статистических данных известно. что вероятность заболеть гриппом во время эпидемии для каждого индивида равняется 0.1. Какая вероятность того, что из 100 проверенных лиц больными окажутся от 20 до 50?

1.108. В магазин зашло восемь покупателей. Вероятность того, что любой из них не уйдёт из магазина без покупки, равняется 0.4.

а) Найти вероятность того, что трое in них что-то купят.

б) Какая вероятность того, что ни одни из них ничего не купит?

1.109. Найти вероятность того, что среди 100 прохожих будет не более 40 брюнетов, если около 30 % населения — брюнеты.

1.110. В процессе производства вероятность дефектов в каждой партии продукции составляет 0.1. Какая вероятность того, что in десяти партий дефекты будут иметь меньше двух партий?

1.111. Найти вероятность того, что среди 1000 новорожденных будет 480 девочек, если вероятность рождения мальчика равняется 0.515.

1.112. Предположим, что вероятности родиться в любой из дней года одинаковые. Найти вероятность того, что среди 500 учеников школы:

а) трое родились 8-го марта:

б) ни одни не родился 1-го января.

1.113. Вероятность попадании в самолёт из винтовки при каждом выстреле равняется 0,001. Производится 3000 выстрелов. Найти вероятность того, что будет хотя бы одно попадание.

1.114. Среди пяти студентов проводится психологический тест на определение типа характера человека. Вероятность того, что для каждого из них будет правильно определён тип характера по результатам тестирования равняется 0,9. Найти вероятность того, что только для трех протестированных студентов будет правильно определён тин характера.

1.115. Среди семян пшеницы 0,6 % семян сорняков. Какова вероятность при случайном отборе 1000 семян обнаружить а) не менее 3 семян сорняков, б) не более 16 семян сорняков, в) ровно 6 семян сорняков?

1.116. Вероятность изготовления стандартной детали на автомате равна 0,95. Изготовлена партия в 20С деталей. Найти наиболее вероятное число нестандартных деталей в этой партии Найти вероятность этого количества нестандартных деталей.

1.117. В камере хранения ручного багажа 80 % всей клади составляют чемоданы, которые вперемешку с другими вещами хранятся на стеллажах. Через окно выдачи были получены все вещи с одного из стелажей в количестве 50 мест. Найти вероятность того, что среди выданных вещей было 38 чемоданов.

1.118. На факультете 730 студентов Вероятность рождения каждого студента в данный день равна $1/365$. Найти вероятность того, что найдутся три студента с одним и тем же днем рождения. Ответ. $P \approx 0{,}18$.

1.119. Вероятность того, что телевизор выдержит гарантийный срок работы, равна 0,8. Найти границы, в которых с вероятностью 0,9955 заключено число телевизоров, выдержавших гарантийный срок службы из 1000 выпущенных.

1.120. В каждом испытании некоторое событие A происходит с вероятностью р=0,5. Произведено 1600 независимых испытаний. Найти границы для частности, симметричные относительно р, которые можно гарантировать с вероятностью 0,95.

1.121. Вероятность наступления события в каждом из независимых испытаний постоянна и равна 0,6. Сколько испытаний необходимо произвести, чтобы вероятность отклонения частоты от 0,6 в ту и другую сторону менее чем на 0,01 была равна 0,995?

1.122. Из группы 200 человек, сдававших экзамен на получение водительских прав в ГАИ 45 человек экзамен не сдало. Оценить вероятности провала и успеха на экзамене. Используя интегральную теорему Лапласа построить приближенные доверительные границы для вероятности успешной сдачи экзамена при 0.8.

1.123. Шестигранную кость подбрасывают 10000 раз. Оценить вероятность отклонения частоты появления шести очков от вероятности появления того же числа очков меньше чем на 0,01.

1.124. Найти число бросаний монеты, при котором с вероятностью 0,7698 можно ожидать, что относительная частота появления орла отклонится от вероятности его появления по абсолютной величине не более чем на 0,02.

1.125. В магазине продаются 5 отечественных и 3 импортных телевизора. Составить закон распределения случайной величины — числа импортных из 4 наудачу взятых телевизоров. Найти функцию распределения и построить ее график.

1.126. Случайная величина X задана функцией распределения

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

Найти вероятность того, что в результате испытании величина X примет значение, заключенное в интервале (0, 1/3).

1.127. Случайная величина X задана на всей оси Ox функцией распределении $F(x)=1/2+(arctgx)/\pi$. Найти вероятность того, что в результате испытания величина X примет значение, заключенное и интервале (0,1).

1.128. Функция распределения непрерывной случайной величины X (времени безотказной работы некоторого устройства) равна $F(x)=1-e^{-x/T}$ $(x\ge 0)$. Найти вероятность безотказной работы устройства за время $x\ge 0$.

1.129. Случайная величина X задания функцией распределения

$$F(x)=\begin{cases} 0 \text{ } при \text{ } x\le 2 \\ 0.5x \text{ } при \text{ } 2<x\le 4 \\ 1 \text{ } при \text{ } x>4 \end{cases}$$

Найти вероятность того, что в результате испытания X примет значение: а) меньшее 0,2; б) меньшее трех; в) не меньшее трех; г) не меньшее пяти.

1.130. Случайная величина X задана функцией распределения

$$F(x)=\begin{cases} 0 \text{ } при \text{ } x\le 0 \\ x^2 \text{ } при \text{ } 0<x\le 1 \\ 1 \text{ } при \text{ } x>1 \end{cases}$$

Найти вероятность того, что в результате четырех независимых испытаний величина X ровно три раза примет значение, принадлежащей интервалу (0,25, 0,75).

1.131. Задана плотность распределения $f(x)$ случайной величины X, возможные значения которой заключены в интервале $(-\infty;+\infty)$. Найти плотность распределения $g(x)$ случайной величины Y, если а) $Y=X^2$ б) $Y=e^{-X^2}$ в) $Y=|X|$ г) $Y=arctgX$ д) $Y=\dfrac{1}{1+X^2}$

1.132. Задача плотность распределения $f(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Нормально распределенной случайной величины X. Найти плотность распределения случайной величины $Y=\dfrac{1}{2}X^2$.

1.133. Дана функций распределений непрерывной случайность величины X:

$$F(x)=\begin{cases} 0 \quad при\ x \le 0 \\ \sin 2x\ при\ 0 < x \le \pi/4 \\ 1\ при\ x > \pi/2 \end{cases}$$

Найти плотность распределения $f(x)$.

1.134. Непрерывная случайная величина X задана плотностью распределения $f(x)=(3/2)\sin 3x$ в интервале $(0,\pi/3)$; вне этого интервала $f(x)=0$. Найти вероятность того, что X примет значение, принадлежащее интервалу $(\pi/6,\pi/4)$.

1.135. Задана плотность распределения непрерывной случайной величины *X:*

$$f(x)=\begin{cases} 0\ при\ x \le 0 \\ \sin x\ при\ 0 < x \le \pi/2 \\ 0\ при\ x > \pi/2 \end{cases}$$

Найти функцию распределения $F(x)$.

1.136. Задана плотность распределения непрерывной случайной величины X:

$$f(x)=\begin{cases} 0\ при\ x \le 1 \\ x-1/2\ при\ 1 < x \le 2 \\ 0\ при\ x > 2 \end{cases}$$

Найти функцию распределения $F(x)$.

1.137. Задана плотность распределения непрерывной случайной величины X:

$$f(x)=\begin{cases} 0 \quad при\ x \le \pi/6 \\ 3\sin 3x \quad при\ \pi/6 < x \le \pi/3 \\ 0 \quad при\ x > \pi/3 \end{cases}$$

Найти функцию распределения $F(x)$.

1.138. Случайная величина X задана плотностью распределения $f(x)=(1/2)x$ в интервале $(0;2)$; вне этого интервала $f(x)=0$. Найти математическое ожидание величины X.

1.139. Случайная величина X в интервале $(-c;c)$; задана плотностью распределения $f(x)=1/\left(\pi\sqrt{c^2-x^2}\right)$; вне этого ни тер вала $f(x)=0$. Найти математическое ожидание величины X.

1.140. Определить математическое ожидание случайной величины x числа попаданий при трёх выстрелах, если вероятность попадания при каждом выстреле P = 0,4.

Подсказка: вероятность значений случайной величины найти по формуле Бернулли.

1.141. Дискретная случайная величина X принимает лишь два значения. Большее из значений 3 она принимает с вероятностью 0,4. Кроме того, известна дисперсия случайной величины D(X) = 6. Найти математическое ожидание случайной величины.

1.142. Случайная величина X задана плотностью распределения $f(x) = c(x^2 + 2x)$ в интервале $(0,1)$; вне этого интервала $f(x) = 0$. Найти: а) параметр c; б) математическое ожидание величины X.

1.143. Найти математическое ожидание случайной величины X, заданной функцией распределения

$$F(x) = \begin{cases} 0 \text{ } при \text{ } x \le 0 \\ x/4 \text{ } при \text{ } 0 < x \le 4 \\ 1 \text{ } при \text{ } x > 4 \end{cases}$$

1.144. Случайная величина X задана плотностью распределения $f(x) = x + 0.5$ в интервале $(0,1)$; вне этого интервала $f(x) = 0$. Найти математическое ожидание функция $Y = X^3$ (не находя предварительно плотности распределения Y).

1.145. Случайная величина X задана плотностью распределения $f(x) = 2\cos 2x$ в интервале $(0, \pi/4)$; вне этого интервала $f(x) = 0$. Найти! а) моду; б) медиану X.

1.146. Поступающий в институт должен сдать 3 экзамена Вероятность сдачи первого экзамена 0.9, второго — 0,8, третьего — 0.7 Следующий экзамен поступающий сдает только в случае успешной сдачи предыдущего. Составить закон распределения числа приходов на экзамен для лица, поступающего в институт. Найти математическое ожидание случайной величины.

1.147. В билете три задачи. Вероятность правильного решения первой задачи равно 0,9, второй -0,8, третье-0,7. Составить закон распределения числа правильно решенных задач в билете и вычислить математическое ожидание и дисперсию.

1.148. В городе 4 коммерческих банка. У каждого риск банкротства в течение года составляет 10% . Составить закон распределения числа банков, которые могут обанкротиться в течение следующего года и найти числовые характеристика этого распределения.

1.149. Вероятность поражения земляники вирусным заболеванием равна 0,2. Составить закон распределения числа кустов земляники, зараженных вирусом, из четырех Посаженных кустов. Найти математическое ожидание и дисперсию этой случайной величины.

1.150. В урне находятся шары трех весов 3,4 и 5 кг соответствующими вероятностями 0,2;0,3;0. Извлекаются два шара с возвращением обратно. Составить закон распределения суммарного веса два извлеченных шаров. Найти математическое ожидание и дисперсию этой случайной величины.

1.151. Найти закон распределения числа пакетов трех акций, по которым владельцем будет получен доход, если вероятность получения дохода по каждому из них равно соответственно 0,5.0,6.0,7.Найти математическое ожидание и дисперсию этой случайной величины.

1.152. В лотерее разыгрывается один автомобиль стоимостью 5000 ден. ед., четыре телевизора стоимостью 250 ден. ед. каждый, пятъ магнитофонов стоимостью 200 ден. ед. каждый. Продано 1000 билетов стоимостью 7 ден. ед. каждый. Составить закон распределения случайной величины *X*-чистого выигрыша, полученного участником лотереи, купившим один билет.

1.153. Дискетная случайная величина X может принимать только два значения x_1 и x_2, причем $x_1 < x_2$. Известны вероятность $p_1 = 0{,}1$ возможного значения x_1, математическое ожидание $M(x)=3,9$. И дисперсия $D(x)=0,09$. Найти закон распределения этой случайной величины.

1.154. Имеется шесть ключей, из которых только один подходит к замку. Составить закон распределения числа попыток при открывании замка, если испробованный ключ в последующих опробованиях участвует. Найти математическое ожидание и дисперсию этой случайной величины.

1.155. В урне 6 белых и 4 чёрных шара. Из урны вынимают 3 шара. Число белых шаров среди вынутых шаров является дискретной случайной величиной X. Найти математическое ожидание и дисперсию этой случайной величины.

1.156. Случайная величина X в интервале $(0;\pi)$ задана плотностью распределения $f(x)=(1/2)\sin x$; вне этого интервала $f(x)=0$. Найти дисперсию X.

1.157. Случайная величина X в интервале $(0,\ 5)$ задана плотностью распределения $f(x)=(2/25)x$; вне этого интервала $f(x)=0$. Найти дисперсию *X*.

1.158. Найти дисперсию случайной величины *X*, заданной функцией распределения

$$f(x)=\begin{cases} 0\,при\,x\le -2 \\ x/4+1/2\ при -2<x\le 2 \\ 1\,при\ \ x>2. \end{cases}$$

1.159. Цена деления шкалы измерительного прибора равна 0,1. Показания прибора округляют до ближайшего целого деления. Полагая, что при отсчете ошибка округления распределена по равномерному закону, найти: 1) математическое ожидание, дисперсию и среднее квадратическое отклонение этой случайной величины; 2) вероятность того, что ошибка округления; а) меньше 0,01; б) больше 0,03.

1.160. Автобусы некоторого маршрута идут строго по расписанию. Интервал движения 4 мин. Найти вероятность того, что пассажир, подошедший к остановке, будет ожидать очередной автобус менее 2 мин.

1.161. Минутная стрелка часов перемещается скачком в конце каждой минуты. Найти вероятность того, что в данное мгновение часы покажут время, которое отличается от истинного не более чем на 10 с.

1.162. Диаметр круга X измерен приближенно, причем $5 \le x \le 6$. Рассматривая диаметр как случайную величину X, распределенную равномерно в интервале $(5; 6)$, найти математическое ожидание и дисперсию площади круга.

1.163. Мастер, осуществляющий ремонт на дому, может появиться в любое время с 10 до 18 часов. Клиент, прождав до 14 часов, отлучился на 1 час. Какова вероятность, что мастер (приход его обязателен) не застанет его дома?

1.164. Владелец антикварного аукциона полагает, что предложение цены за определенную картину будет равномерно распределенной случайной величиной в интервале от 500 тысяча до 2 млн рублей Найти: а) плотность вероятности ; б) вероятность того, что картина будет продана за цену, меньшую чем 675 тыс. ; в) вероятность того, что цена картины будет выше 2 млн рублей.

1.165. Очень наблюдательный, занимающийся кражей предметов искусства вор, который, вероятно, знает хорошо статистику, заметил, что частота, с которой охранники обходят музей, равномерно распрелена между 15 и 60 минутами. Пусть случайная величина X — время (в минутах) до появления охраны. Найти: а) вероятность того, что охранник появится в течение 35 минут после появления вора; б) вероятность того, что охрана не появится в течение 30 минут; в) вероятность того, что охрана появится между 35 и 45 минутами после прихода вора; г) функцию распределения $F(X)$.

1.166. На перекрестке дорог движение регулируется автоматическим светофором, включающим зеленый свет через каждые 2 минуты. Время простоя у этого светофора автомобиля, остановившегося на красный свет, есть случайная величина, распределенная равномерно на интервале $(0; 2)$

минут. Найти среднее время простоя и среднее квадратическое отклонение.

1.167. Коробки с конфетами упаковываются автоматически. Их средняя масса равна 540 г. Известно, что 5 % коробок имеет массу, меньшую 500 г Каков процент коробок, масса которых: а) менее 470 г, 6) от 500 до 550 г, в) более 550 г; г) отличается от средней не более, чем на 30 г (по абсолютной величине)?

1.168. Случайная величина X имеет нормальное распределение с математическим ожиданием $A = 25$. Вероятность попадания X в интервал $(10;15)$ равна 0,09. Чему равна вероятность попадания X в интервал: $a)(35;40)$; $б)(30;30)$ $в)\sigma = 10$?

1.169. Вес пойманной рыбы подчиняется нормальному закону с параметрами $A = 375\,г, \sigma = 25\,г.$ Найти вероятность того, что вес одной рыбы будет: а) от 300 до 425 г; б) не более 450 г, в) больше 300 г.

1.170. Случайная величина X распределена по нормальному закону с $M(X) = 10,\ D(X) = 4$ Найти: $а) P(12 < X < 14);\ б) P(8 < X < 12)$?.

1.171. Случайная величина X имеет нормальное распределение с $A = 0, \sigma = 1.$ Что больше $P(-0,5 \le x \le -0,1)$ или $P(1 \le X \le 2)$?

1.172. Производится взвешивание некоторого вещества без систематических погрешностей Случайные погрешности взвешивания подчинены нормальному закону со средним квадратичным отклонением $\sigma = 20\,Г$. Найти вероятность того, что взвешивание будет произведено с погрешностью, не превосходящей по абсолютной величена 10 г.

1.173. Случайная величина X — ошибки измерений — распределена нормально. Найти вероятность того, что X примет значение между -3σ и 3σ (предполагается, что систематические погрешности отсутствуют).

1.174. Математическое ожидание нормально распределенной случайной величины X равно $a = 3$ и среднее квадратическое отклонение $\sigma = 2$. Написать плотность вероятности X.

1.175. Написать плотность вероятности нормально распределенной случайной величины *X*, зная, что $M(X)=3; D(X)=16.$

1.176. Математическое ожидание и среднее квадратическое отклонение нормально распределенной случайной величины *X* соответственно равны 10 и 2. Найти вероятность того, что в результате испытания *X* примет значение, заключенное в интервале $(12; 14)$.

1.177. Составить закон распределения случайной величины *X*, Записать функцию распределения, построить её график. Вычислить числовые характеристики $M(X)$, $D(X)$, $\sigma(X)$.

В партии 10% бракованных изделий. Наудачу отобрано 5 изделий. *X* - число бракованных изделий среди отобранных. Дискретная случайная величина *X* распределена по биномиальному закону.

1.178. Стрелок производит 4 выстрела по мишени. Вероятность попадания в мишень при каждом выстреле равна 0,3. За каждое попадание стрелку засчитывается 10 очков, Найти закон распределения числа засчитанных очков.

1.179. Опыт состоит из трех независимых подбрасываний одновременно трех монет, каждая из которых с одинаковой вероятностью падает гербом или цифрой вверх. Построить ряд распределения, найти функцию распределения, математическое ожидание и среднее квадратичное отклонение числа одновременного выпадения двух гербов.

Найти вероятность того, что два герба одновременно выпадут хотя бы один раз.

1.180. ОТК проверяет изделия на стандартность, Вероятность того, что изделие стандартно, равна 0,7. Проверено 20 изделий. Найти закон распределения случайной величины X - числа стандартных изделий среди проверенных. Вычислить математическое ожидание, дисперсию и среднеквадратическое отклонение случайной величины.

1.181. По многолетним статистическим данным известно, что вероятность рождения мальчика равна 0,515. Составить закон распределения случайной величины X - числа мальчиков в семье с 4 детьми, Найти математическое ожидание и дисперсию этой случайной величины.

1.182. Производится три независимых опыта, в каждом из которых событие А появляется с вероятностью 0,6. Построить ряд распределения и функцию распределения случайной величины X числа появления события А в трех опытах. Найти числовые характеристики этой случайной величины X.

1.183. Среднее число самолетов, взлетающих с полевого аэродрома за одни сутки, равно 10. Найти вероятность того, что за 6 часов взлетят:

А) три самолета,

Б) не менее двух самолетов.

1.184. На автовокзале время прибытия автобусов различных рейсов объявляет дежурный. Появление информации о различных рейсах происходит случайной и независимо друг от друга. В среднем на автовокзал прибывает 5 рейсов каждые полчаса.

Составьте ряд распределения числа сообщений о прибытии автобусов в течение получаса.

Б) Найдите числовые характеристики этого распределения. Запишите функцию распределения вероятностей и постройте ее график.

Г) Чему равна вероятность того, что в течение получаса прибудут не менее трех автобусов?

Д) Чему равна вероятность того, что в течение четверти часа не прибудет ни один автобус?

1.185. АТС получает в среднем за час 480 вызовов. Определить вероятность того, что за данную минуту она получит: ровно 3 вызова; от 2 до 5 вызовов.

1.186. Случайная величина X распределена по закону Пуассона с параметром λ = 0,8, Необходимо:

выписать формулу для вычисления вероятности $P(X = m)$;

Б) найти вероятность $F(1 < X < 3)$;

В) найти математическое ожидание $M\{2X + 5)$ и дисперсию $D(5 — 2X)$.

1.187. Среднее число ошибочных соединений, приходящееся на одного телефонного абонента в единицу времени, равно 8. Какова вероятность того, что для данного абонента число ошибочных соединений будет больше 4?

1.188. В среднем в магазин заходят 3 человека в минуту. Найти вероятность того, что за 2 минуты в магазин зайдет не более 1 человека.

1.189. Среднее время безотказной работы прибора равно 85 ч. Полагая, что время безотказной работы прибора имеет показательный закон распределения, найти: а) выражение его плотности вероятное функции распределения; б) вероятность того, что в течение 100 ч прибор не выйдет из строя.

1.190. Написать плотность и функцию распределения показательного распределения показательное закона, если параметр $\lambda = 6$.

1.191. Найти параметр λ показательного распределения

А) заданного плотностью $f(x) = 0$ при $x < 0$, $f(x) = 2e^{-2x}$ при $x \geq 0$; б) Заданного функцией распределения $F(x) = 0$ при $x < 0$ и $F(x) = 1 - e^{-0,4x}$ при $x \geq 0$;

1.192. Производится испытание трех элементов, работающих независимо один от другого длительность времени без отказной работа элементов распределена по показательному закону для первого элемента $p_1(t) = 0,1e^{-0,1t}$ для второго - $p_2(t) = 0,2e^{-0,2t}$; для третьего элемента $p_3(t) = 0,3e^{-0,3t}$; Найти вероятность того, что в интервале времени $(0;10)$ часов откажут: а) только один элемент б) только два элемента в) хотя бы один элемент г) все три элемента д) не мне двух элементов.

1.193. $P\%-$ м ресурсом элемента назевается такое число T, что за время T элемент не выходит из строя с вероятностью P. Считается время T непрерывной работы электрической лампочки распределено по показательному закону. Найти вероятность того, что лампочка будут гореть в течение 2 лет, если ее 90% -й ресурс составляет 6 мес.

1.194. Срок службы жесткого диска компьютера – случайная величина, подчиняющаяся экспоненциальному распределению со средней в 12000 часов. Найти долю жестких дисков, срок службы которых превысит

20000 часов.

1.195. Срок службы батареек для слуховых аппаратов приблизительно подчиняется экспоненциальному закону с $\lambda = 1/2$. Какова доля батареек со сроком службы больше чем 9 дней?

1.196. Задача 8. Служащий рекламного агентства утверждает, что время, в течение которого телезрители понята содержание коммерческого рекламного ролика, подчиняется экспоненциальному закону с $\lambda = 0,25$. Дня. Найти долю зрителей, способных вспомнить рекламу спустя 7 дней.

1.197. Используя неравенство Чебышева, оценить вероятность того, что $|\xi - M(\xi)| < 0,3$, если $D(\xi) = 0,0025$. Оценить вероятность, этого же события, если известно, что ξ имеет нормальное распределение.

Если предположить, что ξ имеет нормальное распределение, то $P\{|\xi - M(\xi)| < 0,3\} \approx 1$ (с точностью до 0,0001).

1.198. Масса деталей, изготавливаемых на станке, является случайной величиной, среднее значение которой (математическое ожидание) равняется 1.2 кг. Дисперсия этой величины равняется 0,012. Используя неравенство Чебышева. Оценить вероятность того, что:

а) отклонение массы детали от её среднего значения по абсолютной величине не превысит 0,2;

б) масса детали примет значение от 1,18 до 1,22.

Ответ а) $P \geq 0,7;\ б) P \geq 0.$

1.199. Прибор состоит из 10 элементов, которые работают независимо. Вероятность отказа каждого элемента за время t равняется 0.05. Используя неравенство Чебишева, оценить вероятность того, что абсолютная величина разности между количеством элементов. которые отказали, и средним количеством (математическим ожиданием) отказов за время t будет меньше 2.

1.200. Из 1000 изделий наугад отобраны 200. После проверки среди них выявлено 15 бракованных. Приняв долю бракованных изделий как вероятность изготовления брака, оценить вероятность того, что во всей партии бракованных изделий будет не больше 10 % и не меньше 5%.

1.201. В стране N проживает 50 млн. людей. ВПЧ- тестирование прошли 1000 случайно отобранных граждан, среди которых оказалось 20 инфицированных. Приняв долю инфицированных в исследованной группе как вероятность того, что случайно отобранный гражданин страны N инфицирован, оценить вероятность того, что в данной стране процент ВИЧ-инфицированных составляет не больше 3% и не меньше 1%.

1.202. Средний вес клубня картофеля равен 150 г. Оценить вероятность того, что наудачу взятый клубень картофеля весит не более 500 г?

1.203. Среднее значение скорости ветра у земли в данном пункте равно 16 км/ч. Оценить вероятность того, что в этом пункте скорость ветра (при одном наблюдении) не превысит 80 км/ч.

1.204. Среднее потребление электроэнергии за май населением одного из микрорайонов Минска равно 360000 кВт/ч. Оценить вероятность того, что потребление электроэнергии в мае текущего года превзойдет 1000 000 кВт/ч.

1.205. Среднее квадратическое отклонение ошибки измерения курса самолета $\sigma = 2^{\circ}$, Считая математическое ожидание ошибки измерения равным нулю, оценить вероятность того, что ошибка при данном измерении курса самолета будет более $\sigma = 5^{\circ}$.

1.206. Среднее квадратическое отклонение ошибки измерения азимута равно $30'$ (математическое ожидание равно нулю). Оценить вероятность того, что ошибка среднего арифметического трек независимых измерений не превзойдет 1°.

1.207. Длина изготавливаемых деталей является случайной величиной, среднее значение которой 50 мм. Среднее квадратическое отклонение этой величины равно 0,2 мм. Оценить вероятность того, что отклонение длины изготовленной детали от ее среднего значения по абсолютной величине не превзойдет 0,4 мм.

1.208. За значение некоторой величины принимают среднее арифметическое достаточно большого числа ее измерений. Предполагая, что среднее квадратическое отклонение возможных результатов каждого измерения не превосходит 5 мм, оценить вероятность того, что при 1000 измерений неизвестной величины отклонение принятого значения от истинного по абсолютной величине не превзойдет 0,5 мм.

Вопросы для повторения первой главы.

1.1.На какие классы делятся законы природы и общества по форме проявления причинных связей?

1.2.На какие виды можно подразделить события?

1.3.Что является предметом теории вероятностей?

1.4. Каково значение теории вероятностей для экономических задач?

1.5.Что называется пространством элементарных событий?

1.6.В чем сущность статистического определения вероятности?

1.7.Каково геометрическое определение вероятности?

1.8. Какие события называются несовместными, а какие совместными?

1.9. Что такое противоположное событие и как оно обозначается?

1.10. Какие события называются независимыми, а какие зависимыми?

1.11. Дать определение условной вероятности.

1.12.Чему равна вероятность противоположного события?

1.13. Какие события образуют полную группу событий?

1.14. Напишите формулу полной вероятности и как она выводится?

1.15. Напишите формулу Байеса.

1.16. Что называется схемой Бернулли ?

1.17. Приведите локальную теорему Лапласа?

1.18. О чем идет речь в интегральной теореме Лапласа?

1.19. Дать определение дискретной случайной величине?

1.20. Что такое непрерывная случайная величина?

1.21.Биномиальный закон распределения.

1.22. Каковы особенности геометрического закона распределения ?

1.23. В каких случаях используют распределение Пуассона ?

1.24. Что называется числовыми характеристиками случайной величины.

1.25. Что такое математическое ожидание и как оно определяется?

1.26. Какие случайные величины называются независимыми и что является произведением независимых случайных величин?

1.27. Дать определение дисперсии.

1.28. Среднее квадратическое отклонение и как оно определяется?

1.29. Что называется функцией распределения случайной величины?

1.30. Свойства функции распределения.

1.31. Какими свойствами обладают графики функций распределения непрерывной и дискретной случайной величин?

1.32. Что называется функцией плотности непрерывной случайной величины.

1.33. Как найти функцию распределения, зная функцию плотности распределения.

1.34. Математическое ожидание непрерывной случайной величины?

1.35. Дисперсия непрерывной случайной величины и как она вычисляется?

1.36. Что называется нормальным распределением?

1.37. Как вычисляется математическое ожидание и дисперсия равномерной случайной величины?

1.38. Что называется показательным распределением?

1.39. Как находится вероятность попадания показательной случайной величины в заданный интервал?

1.40. Неравенство Чебышева.

1.41. Что утверждает закон больших чисел в форме Чебышева?

1.42. В чем сущность закона больших чисел и каково его практическое значение?

1.43. Что утверждает закон больших чисел в форме Бернулли?

1.44. Центральная предельная теорема.

Глава 2. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

2.1. Введение в математическую статистику

2.1.1. Основные понятия математической статистики

Математическая статистика – это раздел математики, изучающий методы сбора, систематизации и обработки результатов наблюдений массовых случайных явлений с целью выявления существующих закономерностей. Выводы о закономерностях, которым подчиняются явления, изучаемые методами математической статистики, всегда основываются на ограниченном, выборочном числе наблюдений. Поэтому естественно предположить, что эти выводы при большом числе наблюдений могут оказаться иными. Чтобы быть в состоянии высказать более определённое суждение об изучаемом явлении, математическая статистика опирается на теорию вероятностей.

Оценив неизвестные величины или зависимости между ними по данным наблюдений, исследователь выдвигает ряд гипотез, предположений о том, что рассматриваемое явление можно описать той или иной вероятностной теоретической моделью. Далее, используя математико-статистические методы, можно дать ответна вопрос, какую из гипотез или моделей следует принять. Именно эта модель и есть закономерность изучаемого явления.

Основные задачи математической статистики состоят в разработке методов:

1) организации и планирования статистических наблюдений;
2) сбора статистических данных;
3) «свертка информации», то есть методов группировки и сокращения статистических данных с целью сведения большого числа таких данных к небольшому числу параметров, которые в сжатом виде характеризуют всю исследуемую совокупность;
4) анализа статистических данных;
5) принятия решений, рекомендаций и выводов на основе анализа статистических данных;
6) прогнозирования случайных явлений.

Одним из основных методов статистического наблюдения является выборочный метод. Рассмотрим основные понятия этого метода.

2.1.2. Вариационные ряды

Генеральной совокупностью называется совокупность всех мыслимых наблюдений, которые могли бы быть сделаны при данном реальном комплексе условий измерений. Число членов, образующих генеральную совокупность, называется *объёмом* генеральной совокупности.

Выборочной совокупностью или просто *выборкой объёма* n называется совокупность n объектов, отобранных из исследуемой генеральной совокупности.

Вариантами называются различные значения признака, а *варьированием* – изменение значений признака.

Расположение наблюдаемых значений в порядке возрастания или убывания называется *ранжированием* выборки.

Если признак по своей сущности таков, что различные значения не могут отличаться друг от друга меньше чем на некоторую конечную величину, то говорят, что это *дискретно варьирующийся признак*.

Число, показывающее, сколько раз встречается вариант x в ряде наблюдений, называется *частотой* варианта m_x.

Вместо частоты варианта x можно рассматривать её отношение к общему числу наблюдений n, которое называется *частостью* варианта x и обозначается ω_x. Так как общее число наблюдений равно сумме частот всех вариантов $n = \sum m_x$, то справедлива следующая цепочка равенств:

$$\omega_x = \frac{m_x}{n} = \frac{m_x}{\sum m_x}.$$

Таблица, позволяющая судить о распределении частот (или частостей) между вариантами, называется *дискретным вариационным рядом*.

Наряду с понятием частоты используется понятие *накопленной частоты*, которую обозначают $m_x^{\text{нак}}$.

Скольких наблюдениях признак принял значения, меньшие заданного значения x. Отношение накопленной частоты к общему числу наблюдений называют *накопленной частостью* и обозначают $\omega_x^{\text{нак}}$.

Признак называется *непрерывно* варьирующим, если он может принять любое значение в некотором числовом интервале.

По полученным данным такого признака трудно выявить характерные черты варьирования значений признака. Построение дискретного вариационного ряда также не даст желаемых результатов (слишком велико число вариантов). Для получения ясной картины нужно объединить полученные значения в несколько интервалов.

Таблицу, позволяющую судить о распределении частот (или частостей) между интервалами варьирования значений признака, называют *интервальным вариационным рядом*.

Интервальный вариационный ряд строят по данным наблюдений за непрерывно варьирующим признаком, а также за дискретно варьирующим, если велико число наблюдавшихся вариантов. Дискретный вариационный ряд строят только для дискретно варьирующего признака.

Для построения интервального вариационного ряда необходимо определить величину интервала, установить полную шкалу интервалов, в соответствии с ней сгруппировать результаты наблюдений. Для определения оптимального интервала h, то есть такого, при котором построенный интервальный ряд не был бы слишком громоздким и в то же время позволял выявить характерные черты рассматриваемого явления, можно использовать *формулу Стэрджеса:*

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \lg n},$$

где $x_{\max}$ и $x_{\min}$ – соответственно максимальный и минимальный варианты.

Величина равная $(1+3{,}322\lg n)$ характеризует число интервалов, так как количество интервалов должно быть целым числом, то её округляют до целого значения. При этом желательно, чтобы было минимальным для возможных значений m. Длина интервала имеет ту же точность, что и варианты.

Пример 79. Найти оптимальную длину интервала по следующим данным: $x_{\max}=74$, $x_{\min}=16$, $n=50$, причём значения вариантов – целые числа.

Решение. Вычислим $(1+3{,}322\lg 50)\approx 6{,}64$. Возможное количество интервалов 7 или 6. Соответствующие им длины интервалов равны

$$h_1=\frac{74-16}{7}\approx 9;\quad h_2=\frac{74-16}{6}\approx 10.$$

Тогда $\Delta_1=9\cdot 7-(74-16)=5$, $\Delta_2=10\cdot 6-(74-16)=2$. Так как $\Delta_1=5>\Delta_2=2$, то для уменьшения погрешностей желательно взять шесть интервалов, имеющих длину 10.

Пример 80. Во время исследования количественного признака X из генералной совокупности была получена выборка

$$4,3,6,4,7,2,5,1,2,5,4,4,3,5,6,3,4,1,3,4.$$

Найти объём выборки, построить вариационный ряд выборки и её статистическое распределение.

Решение: Поскольку выборка состоит из 20 значений, то объём выборки $n=20$.Построим вариационный ряд выборки:

$$1,1,2,2,3,3,3,3,4,4,4,4,4,4,5,5,5,6,6,7$$

В данной выборке всего семь разных значений, т.с. вариант:

$$1,2,3,4,5,6,7.$$

Найдем их частоты:

$n_1=2$: $n_2=2$: $n_3=4$: $n_4=6$: $n_5=3$: $n_6=2$: $n_7=1$:

Запишем искомое статическое распределение:

x_1	1	2	3	4	5	6	7
n_1	2	2	4	6	3	2	1

Пример 81. Выборка задана распределением частот:

x_i	3	5	7	10	15
ω_i	0.1	0.2	0.35	0.2	0.15

Решение: Вычислим объём выборки:

$$n = 2 + 4 + 7 + 4 + 3 = 20.$$

Определим относительный частот:

$$\omega_1 = \frac{2}{20} = 0{,}1 \; \omega_2 = \frac{4}{20} = 0{,}2 \; \omega_3 = \frac{7}{20} = 0{,}35 \; \omega_4 = \frac{4}{20} = 0{,}2 \; \omega_5 = \frac{3}{20} = 0{,}15.$$

Итак, искомое распределение относительных частот имеет вид

Контрол: $0.1 + 0.2 + 0.35 + 0.2 + 0.15 = 1.$

2.1.3. Графическое изображение вариационного ряда

Графическое изображение вариационного ряда позволяет представить в наглядной форме закономерности варьирования значений признака. Наиболее широко используются следующие виды графического изображения вариационных рядов: полигон, гистограмма, кумулятивная кривая.

Полигон, как правило, служит для изображения дискретного вариационного ряда. Для его построения в прямоугольной системе координат наносят точки с координатами $(x_i; m_i)$, где x_i – вариант; m_i – соответствующая ему частота. Иногда вместо точек $(x_i; m_i)$ строят точки $(x_i; \omega_i)$. Затем эти точки последовательно соединяют отрезками в порядке возрастания x_i. Полученная линия носит название полигон частот или полигон относительных частот. Иногда интервальный ряд изображают с помощью полигона. В этом случае интервалы заменяют их серединными значениями и к ним относят интервальные частоты. Для полученного дискретного ряда строят полигон.

Гистограмма служит для изображения только интервального вариационного ряда. Для её построения в прямоугольной системе координат по оси абсцисс откладывают отрезки, изображающие интервалы варьирования, и на них строят прямоугольники с высотами, равными частотам (или частостям) соответствующего интервала. В результате получают ступенчатую фигуру, состоящую из прямоугольников, которую называют гистограммой.

Кумулятивная кривая (кривая накопленных частот или накопленных частостей) строится следующим образом. Если вариационный ряд дискретный, то в прямоугольной системе координат строят точки с координатами $\left(x_i; m_i^{\text{нак}}\right)$, где x_i – вариант; $m_i^{\text{нак}}$ – соответствующая накопленная частота. Полученные точки соединяют отрезками.

Если вариационный ряд интервальный, то по оси абсцисс откладывают интервалы. Верхним границам интервалов соответствуют накопленные частоты (или накопленные частости); нижней границе первого интервала – накопленная частота, равная нулю. Построив кумулятивную кривую, можно приблизительно установить число наблюдений (или их долю в общем количестве наблюдений), в которых признак принял значения, меньшие заданного. Иногда кумулятивную кривую называют *кумулятой*.

Пример 82. Построить полигон частот, гистограмму, кумуляту для интервального вариационного ряда

Интервалы	4–8	8–12	12–16	16–20	20–24	24–28
Частоты	4	6	12	14	8	6

Решение. Построим полигон частот. Вынесем на координатную плоскость точки $(x_i; m_i)$, где x_i – середины соответствующих интервалов:

Интервалы	4–8	8–12	12–16	16–20	20–24	24–28
x_i	6	10	14	18	22	26
m_i	4	6	12	14	8	6

Соединив точки с координатами $(x_i; m_i)$ ломаной кривой, получим полигон частот (Рис. 2).

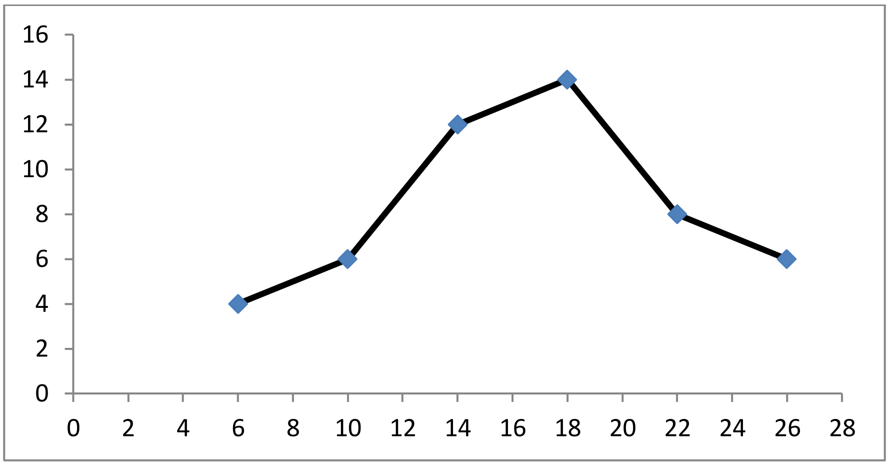

Рис.2. Полигон частот.

Для построения гистограммы отложим на оси абсцисс интервалы варьирования и на них как на основаниях построим прямоугольники с высотами, равными соответствующим частотам (Рис. 3).

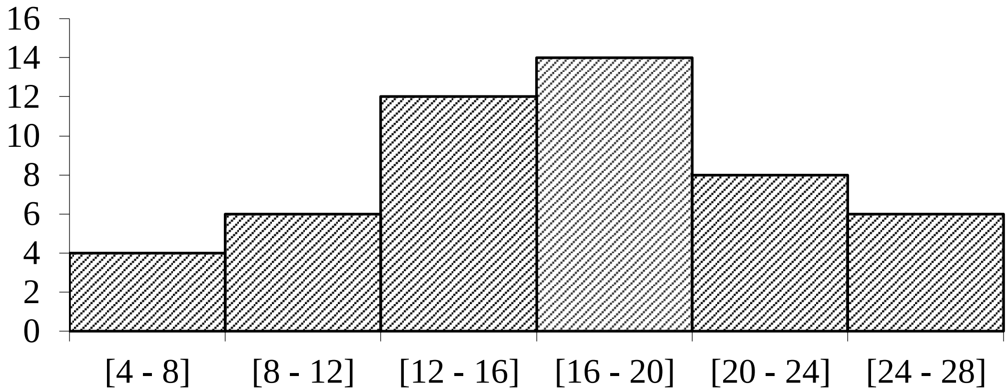

Рис. 3. Кумулят.

Для построения кумулятивной кривой отложим на координатной плоскости точки (4; 0), (8; 4), (12; 10), (16; 22), (20; 36), (24; 44), (28; 45) и соединим их ломаной линией, получим кумуляту (Рис.4).

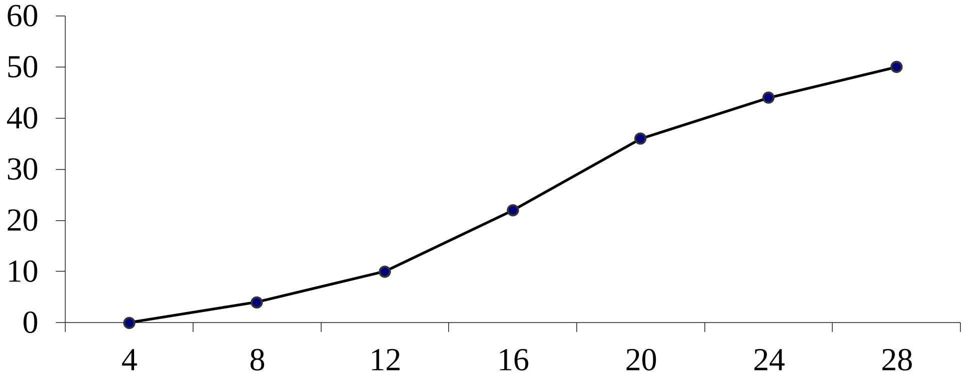

Рис. 4. Ломаной кумулят.

2.1.4. Эмпирическая функция распределения

Эмпирической функцией распределения случайной величины X называют функцию $F^*(x)$, определяющую для каждого значения x частость события $\{X < x\}$: $F^*(x) = \dfrac{n_x}{n}$, где n_x – число вариантов x_i, меньших x; n – объём выборки.

Эмпирическая функция распределения обладает всеми свойствами интегральной функции распределения. Действительно, из определения эмпирической функции распределения следует, что:

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

2) $F^*(x)$ – неубывающая функция;

3) если x_1 – наименьшее, а x_n – наибольшее наблюдённое значение, то $F^*(x) = 0$ при $x \le x_2$ и $F^*(x) = 1$при $x > x_n$.

Основное значение эмпирической функции распределения состоит в том, что она используется в качестве оценки функции распределения $F(x) = p(X < x)$.

Пример 83. Построить эмпирическую функцию распределения для вариационного ряда

Варианты	6	10	14	18
Частота	4	8	6	2

Решение. Согласно определению эмпирическая функция распределения имеет вид:

$$F^*(x) = \begin{cases} 0 & x \le 0; \\ 0{,}2 & 6 < x \le 10; \\ 0{,}5 & 10 < x \le 14; \\ 0{,}8 & 14 < x \le 18; \\ 1 & x > 18. \end{cases}$$

И её график имеет вид (Рис. 5):

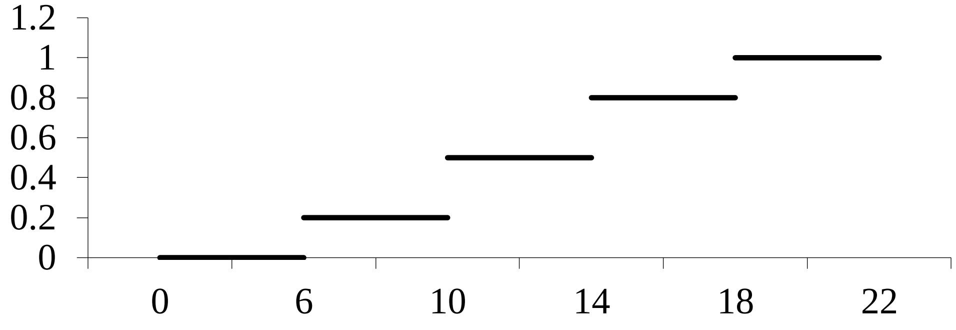

Рис. 5. График эмпирической функций.

Пример 84. Найти эмпирическую функцию распределения по данному распределению выборки:

x_i	3	5	7	10	15
n_i	2	4	7	4	3

Решение. Определим объём выборки: $n = 2+4+7+4+3 = 20.$

Поскольку наименьшая варианта равняется трём,

$$F^*(x) = 0$$

при всех $x \le 3$.

Значение $X < 5$ а именно $x_1 = 3$ наблюдалось дважды поэтому

$$F^*(x) = \frac{2}{20} = 0.1$$

при $3 < x \le 5$.

Значение $X < 7$ а именно $x_1 = 3$ и $x_2 = 5$ наблюдалось 2+4=6 раз, поэтому

$$F^*(x) = \frac{6}{20} = 0.3$$

при $5 < x \le 7$.

Итак запишемм искомую эмперическую функцию:

$$F^*(x) = \begin{cases} 0, & при \quad x \le 3; \\ 0.1 & при \quad 3 < x \le 5; \\ 0.3 & при \quad 5 < x \le 7; \\ 0.65 & при \quad 7 < x \le 10; \\ 0.85 & при \quad 10 < x \le 15; \\ 1 & при \quad x > 15 \end{cases}$$

2.1.5. Средние величины

Средние величины являются как бы «представителями» всего ряда наблюдений, поскольку вокруг них концентрируются наблюдавшиеся значения признака. Заметим, что только для качественно однородных наблюдений имеет смысл вычислять средние величины.

Различают несколько видов средних величин: средняя арифметическая, средняя геометрическая, средняя гармоническая, средняя квадратическая, средняя кубическая и так далее. При выборе вида средней величины необходимо ответить на вопрос: какое свойство ряда мы хотим представить средней величиной или, иначе говоря, какая цель преследуется при вычислении средней? Это свойство, получившее название *определяющего*, и определяет вид средней. Понятие определяющего свойства впервые введено советским статистиком А. Я. Боярским.

Наиболее распространённой средней величиной является средняя арифметическая. Пусть $x_1, x_2, ..., x_n$ – данные наблюдений; $\bar{x}$ – средняя арифметическая. Свойство, определяющее среднюю арифметическую, формулируется следующим образом: *сумма результатов наблюдений должна остаться неизменной, если каждое из них заменить средней арифметической*, то есть $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \bar{x}$. Так как $\bar{x} = \mathrm{const}$, то $\sum_{i=1}^{n} x_i = n \cdot \bar{x}$. Отсюда получаем следующую формулу для вычисления средней арифметической по данным наблюдений:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i .$$

Если по наблюдениям построен вариационный ряд, то средняя арифметическая

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \cdot m_i ,$$

где x_i – вариант, если ряд дискретный, и середина интервала, если ряд интервальный; m_i – соответствующая частота.

Среднюю арифметическую, вычисленную по данным выборки, называют *выборочной средней*.

Очевидно, что если по данным наблюдений построен дискретный вариационный ряд, то обе формулы дают одинаковые значения выборочной средней. Если же по наблюдениям построен интервальный ряд, то выборочные средние, вычисленные по этим формулам, могут не совпадать, так как во второй формуле значения признака внутри каждого интервала принимаются равными центрам интервалов. Ошибка, возникающая в результате такой замены, вообще говоря, очень мала, если наблюдения распределены равномерно вдоль каждого интервала, а не скапливаются к одноимённым границам интервалов.

Основные свойства выборочной средней:

1. Сумма отклонений результатов наблюдений от выборочной средней равна нулю или сумма произведений отклонений вариантов от выборочной средней на соответствующие частоты равна нулю.

2. Если все результаты наблюдений уменьшить (увеличить) на одно и то же число, то и выборочная средняя уменьшится (увеличится) на это же число.

3. Если все результаты наблюдений уменьшить (увеличить) в одно и то же число раз, то выборочная средняя уменьшится (увеличится) во столько же раз.

4. Выборочная средняя алгебраической суммы соответствующих значений признака нескольких рядов наблюдений с одинаковым числом наблюдений равна алгебраической сумме выборочных средних этих рядов.

Пример 85. Найти выборочную среднюю для выборки, представленной интервальным вариационным рядом:

Интервалы	3–7	7–11	11–15	15–19	19–23	23–27
Частоты	6	9	11	12	8	4

Решение. Середины интервалов равны:

5; 9; 13; 17; 21; 25.

Сумма частот равна

$$n = \sum m_i = 6 + 9 + 11 + 12 + 8 + 4 = 50\,.$$

Тогда выборочная средняя равна

$$\bar{x} = \frac{1}{50}\left(5\cdot 6 + 9\cdot 9 + 13\cdot 11 + 17\cdot 12 + 21\cdot 8 + 25\cdot 4\right) = \frac{726}{50} = 14{,}52\,.$$

Пример 86. Задано статистическое распределение выборки:

x_i	1	3	4	7	10	12	15
n_i	5	2	12	7	4	3	2

Найти выборочное среднее выборочную дисперсию и среднее квадратическое отклонение выборки.

Решение. Вычислим объём выборки:

$$n = 5 + 2 + 12 + 7 + 4 + 3 + 2 = 35.$$

Найдём соответственно выборочное среднее выборочную дисперсию и среднее квадратическое отклонение выборки:

$$\bar{x}_n = \frac{\sum_{i=1}^{7} x_i n_i}{n} = \frac{1*5 + 3*2 + 4*12 + 7*7 + 10*4 + 12*3 + 15*2}{35} = \frac{214}{35} \approx 6.1;$$

$$D_n = \frac{\sum_{i=1}^{7} (x_i - \bar{x}_n)^2 n_i}{n} \approx 15.2;$$

$$\sigma = \sqrt{D_n} \approx 3.9$$

Если задано интервальное статистическое распределение ввыборки то выборочное среднее выборочную дисперсию и среднее квадратичское отклонение выборки ищут с помощью такого статического распределения : вариантами считаются середины частных интервалов а частоты или относительное частоты остаются теми же.

Пример 87. Задано интервальное статистическое распределение выборки:

$(x_i; x_{i+1}]$	(0;2]	(2;4]	(4;6]	(6;8]	(8;10]	(10;12]	(12;14]
ω_i	0.1	0.2	0.3	0.1	0.1	0.1	0.1

Найти выборочное среднее выборочную дисперсию и среднее квадратическое отклонение выборки.

Решение. Сначала превратим данное интервальное статистическое распределение выборки и точечное, определив середины частных интервалов:

$$x_1 = \frac{x_1 + x_2}{2} = \frac{0+2}{2} = 1; \quad x_2 = \frac{x_2 + x_3}{2} = \frac{2+4}{2} = 3; \quad x_3 = \frac{x_3 + x_4}{2} = \frac{4+6}{2} = 5;$$

$$x_4 = \frac{x_4 + x_5}{2} = \frac{6+8}{2} = 7; \quad x_5 = \frac{x_5 + x_6}{2} = \frac{8+10}{2} = 9; \quad x_6 = \frac{x_6 + x_7}{2} = \frac{10+12}{2} = 11;$$

$$x_7 = \frac{x_7 + x_8}{2} = \frac{12+14}{2} = 13;$$

Итак, мы получили статистическое распределение выборки:

x_i	1	3	5	7	9	11	13
n_i	0,1	0,2	0,3	0,1	0,1	0,1	0,1

Найдём соответственно выборочное среднее, выборочную дисперсию и среднее квадратическое отклонение выборки:

$$\overline{x}_n = \sum_{i=1}^{7} x_i \omega_i = 1*0{,}1 + 3*0{,}2 + 5*0{,}3 + 7*0{,}1 + 9*0{,}1 + 11*0{,}1 + 13*0{,}1 = 6{,}2$$

$$D_n = \sum_{i-1}^{7} (x_i - \overline{x}_n)^2 \omega_i = 12{,}96; \quad \sigma = \sqrt{D_n} = 3{,}6.$$

2.1.6. Медиана и мода

Наряду со средними величинами в качестве описательных характеристик вариационного ряда применяют медиану и моду.

Медианой $\widetilde{M}_e$ называют значение признака, приходящееся на середину ранжированного ряда наблюдений.

Пусть проведено нечётное число наблюдений, то есть $n = 2q - 1$, и наблюдения про ранжированы и выписаны в следующий ряд: $x_1, x_2, ..., x_{q-1}, x_q, x_{q+1}, ..., x_n$. Здесь x_i – значение признака, занявшее i-е порядковое место в ранжированном ряду. На середину ряда приходится значение x_q. Следовательно, $\widetilde{M}_e = x_q$.

Если проведено чётное число наблюдений, то есть $n = 2q$, то на середину ранжированного ряда $x_1, x_2, ..., x_{q-1}, x_q, x_{q+1},..., x_n$ приходятся значения x_q и x_{q+1}. В этом случае за медиану принимают среднюю арифметическую значений x_q и x_{q+1}, то есть $\tilde{M}_e = \dfrac{x_q + x_{q+1}}{2}$.

Для интервального вариационного ряда медиана определяется по формуле

$$\tilde{M}_e = a_e + h \cdot \frac{\frac{n}{2} - m_e^{\text{нак}}}{m_e},$$

где a_e – начало медианного интервала, то есть такого, которому соответствует первая из накопленных частот, большая или равная половине всех наблюдений; $m_e^{\text{нак}}$ – частота, накопленная к началу медианного интервала; m_e – частота медианного интервала.

Модой $\tilde{M}_0$ называют такое значение признака, которое наблюдалось наибольшее число раз. Нахождение моды для дискретного вариационного ряда не требует каких-либо вычислений, так как ею является вариант, которому соответствует наибольшая частота.

В случае интервального вариационного ряда мода вычисляется по следующей формуле:

$$\tilde{M}_0 = a_0 + h \cdot \frac{m_0 - m_0'}{2m_0 - m_0' - m_0''},$$

где a_0 – начало модального интервала, то есть такого, которому соответствует наибольшая частота; m_0 – частота модального интервала; m_0' – частота интервала, предшествующего модальному; m_0'' – частота интервала, следующего за модальным.

Моду используют в случаях, когда нужно ответить на вопрос, какой товар имеет наибольший спрос, каковы преобладающие в данный момент уровни производительности труда, себестоимости. Модальные производительность и себестоимость помогают вскрыть ресурсы, имеющиеся в экономике.

Пример 88. Вычислить моду и медиану для интервального ряда

Интервалы	3–7	7–11	11–15	15–19	19–23	23–27
Частоты	6	9	11	12	8	4

Решение. Найдём объём выборки по формуле $n=\sum m_i$, или $n=\sum m_i=6+9+11+12+8+4=50$. Вычислим накопленные частоты $m_e^{\text{нак}}=6<25$; $m_2^{\text{нак}}=6+9=15<25$; $m_3^{\text{нак}}=6+9+11=26>25$, следовательно, медианным является интервал $[11;\ 15]$. Тогда $a_e=11$, $m_e^{\text{нак}}=15$, $m_e=11$ и $\tilde{M}_e=11+4\cdot\dfrac{25-15}{11}\approx 14{,}36$. Модальным является интервал $[15;\ 19]$, следовательно, $a_0=15$, $m_0=12$, $m_0'=11$, $m_0''=8$.

Тогда $\tilde{M}_0=15+4\cdot\dfrac{12-11}{24-11-8}=15{,}8$.

Пример 88. На одному отрезков железной дороги планируется организовать остановку пассажирского поезда. Распределение населенных пунктов с численностью их населения приведено в таблице.

На каком километре железной дороги расположен населённый пункт, км	0	2	5	5	8	0	3
Численность населения, тыс. чел.	5	2	3	10	1	4	6

На каком километре железной дороги нужно расположить эту остановку, чтобы суммарное расстояние, которое будут покрывать потенциальные пассажиры до этой остановки было наименьшем.

Решение. Поскольку согласно свойству медианы сумма абсолютных от любой другой величины для решения примера нужно найти медиану.

Сначала определим объём выборки:

$$n = 5 + 2 + 3 + 10 + 1 + 4 + 6 = 31.$$

Итак, серединой (средним членом) вариационного ряд можно записать в виде

$$\underbrace{10,\ldots,10}_{5\,раз},12,12,15,15,15,\underbrace{25,\ldots,25}_{10\,раз},28,\underbrace{30,\ldots,30}_{4\,раз},\underbrace{33,\ldots,33}_{6\,раз}$$

Легко увидеть, что $x_{16} = 25$, т.е. остановку следует расположить на 25-м километре железной дороги.

Пример 89. Задано интервальное статистическое распределение выборки:

$(x_i;x_{i+1}]$	(10;15]	(10;15]	(10;15]	(10;15]	(10;15]	(10;15]	(10;15]
n_i	7	4	5	1	12	3	18

Найти медиану, моду и вариационный размах.

Решение. Определим объём выборки:

$$n = 7 + 4 + 5 + 1 + 12 + 3 + 18 = 50.$$

Медианным частным интервалом будет пятый интервал, поскольку это первый интервал, для которого сумма частот всех предыдущих частных интервалов включительно с данным превышает половину объёма выборки:

$$7 + 4 + 5 + 1 + 12 = 29 > \frac{n}{2} = \frac{50}{2} = 25.$$

Модальнқм частнқм интервалом будет последний интервал. Поскольку он имеет наибольшую частоту.

Найдем начало медианного частного интервала $x_{Me}^{\min}$.длину медианного частного интервала h_{Me}, номер предыдущего к медианному частного интервала $Me-1$, частоту медианного частного интервала n_{Me} начало модального частного интервала $x_{Me}^{\min}$.длину модального частного интервала h_{Mo} частоту модального частного интервала n_{Mo} частоту предыдущего к модального частного интервала n_{Mo-1} частоту следующего за модальным частного интервала n_{Mo+1}, наибольшее наименьшее значения выборки:

$$x_{Me}^{\min} = 30; \qquad h_{Me} = 35 - 30 = 5; \qquad Me - 1 = 4$$

$$n_{Me} = 12; \qquad x_{Me}^{\min}. = 40; \qquad h_{Mo} = 45 - 40 = 5;$$

$$n_{Mo} = 18; \qquad n_{Mo-1} = 3; \qquad n_{Mo+1} = 0;$$

$$x_{max} = 45; \qquad x_{min} = 10;$$

Искомые медиана, мода и вариационный размах соответственно будут такие:

$$Me = x_{Me}^{\min} + h_{Me}\frac{\frac{1}{2}n - \sum_{i=1}^{Me-1} n_i}{n_{Me}} = 30 + 5 \cdot \frac{\frac{1}{2} \cdot 50 - \sum_{i=1}^{4} n_i}{12} = \frac{100}{3} \approx 33.3$$

$$Mo = x_{Mo}^{\min} + h_{Mo}\frac{n_{Mo} - n_{Mo-1}}{2n_{Mo} - n_{Mo-1} + n_{Mo+1}} = \frac{465}{11} \approx 42{,}34$$

$$R = x_{\max} - x_{\min} = 45 - 10 = 35.$$

2.1.7. Показатели вариации

Средние величины, характеризуя вариационный ряд числом, не отражают изменчивости наблюдавшихся значений признака, то есть вариацию. Простейшим показателем вариации является *вариационный размах* R_0, равный разности между наибольшим и наименьшим вариантами:

$$R_0 = x_{\max} - x_{\min}.$$

Вариационный размах – приближённый показатель вариации, так как почти не зависит от изменения вариантов, а крайние варианты, которые используются для его вычисления, как правило, ненадёжны.

Более содержательными являются меры рассеяния наблюдений вокруг средних величин. Выборочная средняя является основным видом средних, поэтому ограничимся рассмотрением мер рассеяния наблюдений вокруг выборочной средней.

Эмпирической дисперсией S^2 называют среднюю арифметическую квадратов отклонений результатов наблюдений от их выборочной средней:

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

Если по результатам наблюдений построен вариационный ряд, то эмпирическая дисперсия

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot m_i .$$

Вместо эмпирической дисперсии в качестве меры рассеяния наблюдений вокруг средней арифметической часто используют эмпирическое среднее квадратическое отклонение, равное арифметическому значению корня квадратного из дисперсии и имеющее ту же размерность, что и значения признака. Для вариационного ряда среднее квадратическое отклонение равно

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot m_i} .$$

Свойства эмпирической дисперсии:

1. Дисперсия постоянной величины равна нулю.

2. Если все результаты наблюдений уменьшить (увеличить) на одно и то же число, то дисперсия не изменится.

3. Если все результаты наблюдений уменьшить (увеличить) в одно и то же число k раз, то дисперсия уменьшится (увеличится) в k^2 раз.

4. Эмпирическая дисперсия равна разности между средней арифметической квадратов наблюдений и квадратом средней арифметической $S^2 = \overline{x^2} - (\bar{x})^2$.

Пример 90. Найти эмпирическую дисперсию и выборочное среднее квадратическое отклонение для следующего интервального вариационного ряда:

Интервалы	3–7	7–11	11–15	15–19	19–23	23–27
Частоты	6	9	11	12	8	4

Решение. Эмпирическую дисперсию найдём по формуле

$$S^2 = \overline{x^2} - (\bar{x})^2;$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\,2} \cdot m_i,$$

где x_i – середины интервалов; m_i – соответствующая частота, тогда

$$\overline{x^2} = \frac{1}{50}\left(5^2 \cdot 6 + 9^2 \cdot 9 + 13^2 \cdot 11 + 17^2 \cdot 12 + 21^2 \cdot 8 + 25^2 \cdot 4\right) =$$

$$= \frac{12234}{50} = 244{,}68;$$

$$\bar{x} = 14{,}52;$$

$$S^2 = 244{,}68 - 14{,}52^2 = 33{,}8496.$$

2.1.8. Эмпирические центральные и начальные моменты

Средняя арифметическая и дисперсия вариационного ряда являются частными случаями более общего понятия о моментах вариационного ряда.

Эмпирическим начальным моментом $\tilde{V}_k$ порядка k называют взвешенную среднюю арифметическую k-х степеней вариантов $\tilde{V}_k = \overline{x^k} = \frac{1}{n}\sum_{i=1}^{n} x_i^k \cdot m_i$.

Эмпирическим центральным моментом $\tilde{\mu}_k$ порядка k называют взвешенную среднюю арифметическую k-х степеней отклонений вариантов от их средней арифметической

$$\tilde{\mu}_k = \overline{\left(x - \bar{x}\right)^k} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^k \cdot m_i.$$

Эмпирический центральный момент нулевого порядка равен

$$\tilde{\mu}_0 = \overline{\left(x - \bar{x}\right)^0} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^0 \cdot m_i = \frac{1}{n} \cdot n = 1.$$

Эмпирический центральный момент первого порядка равен

$$\tilde{\mu}_1 = \overline{\left(x - \bar{x}\right)^1} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^1 \cdot m_i = 0.$$

Эмпирический центральный момент второго порядка равен

$$\tilde{\mu}_2 = \overline{\left(x - \bar{x}\right)^2} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \cdot m_i = \tilde{V}_2 - \tilde{V}_1^2 = S^2.$$

Эмпирический центральный момент третьего порядка равен

$$\tilde{\mu}_3 = \overline{\left(x - \bar{x}\right)^3} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3 \cdot m_i = \tilde{V}_3 - 3\tilde{V}_2 \cdot \tilde{V}_1 + 2\tilde{V}_1^3.$$

Эмпирический центральный момент четвёртого порядка равен

$$\tilde{\mu}_4 = \overline{\left(x - \bar{x}\right)^4} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^4 \cdot m_i = \tilde{V}_4 - 4\tilde{V}_3 \cdot \tilde{V}_1 + 6\tilde{V}_2 \cdot \tilde{V}_1 - 3\tilde{V}_1^4.$$

Рассмотрим свойства центральных моментов, которые позволяют значительно упростить их вычисление:

1. Если все варианты уменьшить (увеличить) на одно и то же число, то центральный момент k-го порядка не изменится.

2. Если все варианты уменьшить (увеличить) в одно и то же число q раз, то центральный момент k-го порядка уменьшится (увеличится) в q^k раз.

2.1.9. Эмпирические асимметрия и эксцесс

Эмпирическим коэффициентом асимметрии $\tilde{A}$ называют отношение центрального момента третьего порядка к кубу среднего квадратического отклонения $\tilde{A} = \dfrac{\tilde{\mu}_3}{S^3} = \dfrac{\tilde{\mu}_3}{\sqrt{\tilde{\mu}_2^3}}$.

Если полигон вариационного ряда скошен, то есть одна из его ветвей, начиная от вершины, зримо длиннее другой, то такой ряд называют *асимметричным*. Если в вариационном ряду преобладают варианты, меньшие $\bar{x}$, то эмпирический коэффициент асимметрии отрицателен, и в этом случае имеет место *левосторонняя асимметрия*. Если же в вариационном ряду преобладают варианты, большие $\bar{x}$, то эмпирический коэффициент асимметрии положителен, и в этом случае имеет место *правосторонняя асимметрия*. При левосторонней асимметрии левая ветвь длиннее правой. При правосторонней более длинной является правая ветвь.

Эмпирический коэффициент асимметрии не имеет ни верхней, ни нижней границы, что снижает его ценность как меры асимметрии. Практически коэффициент асимметрии редко бывает особенно велик, а для умеренно асимметричных рядов он обычно меньше единицы.

Эмпирическим эксцессом или *коэффициентом крутости* $\tilde{E}$ называют уменьшенное на три единицы отношение центрального момента четвёртого порядка к четвёртой степени среднего квадратического отклонения:

$$\tilde{E} = \frac{\tilde{\mu}_4}{S^4} - 3 = \frac{\tilde{\mu}_4}{\tilde{\mu}_2^2} - 3.$$

За стандартное значение эксцесса принимают нуль-эксцесс нормальной кривой распределения. Кривые, у которых эксцесс отрицательный, по сравнению с нормальной менее крутые, имеют более плоскую вершину и называются плосковершинными. Кривые с положительным эксцессом более крутые по сравнению с нормальной кривой, имеют более острую вершину и называются островершинными.

Коэффициенты асимметрии и эксцесса не зависят от выбора начала отсчёта и единицы измерения, то есть для любых постоянных $a \neq 0$ и b выполнены равенства:

$$\tilde{A}(ax+b) = \tilde{A}(x);$$
$$\tilde{E}(ax+b) = \tilde{E}(x).$$

2.2. Методы расчёта основных характеристик вариационного ряда

2.2.1. Метод условных вариантов для расчёта числовых характеристик вариационного ряда

Значения вариантов могут быть достаточно большими, и, следовательно, вычисление числовых характеристик достаточно трудоёмко. Поэтому для дискретного вариационного ряда при вычислении коэффициентов асимметрии и эксцесса желательно перейти к условным вариантам по формуле

$$u_i = \frac{x_i - C}{h},$$

где h – шаг вариационного ряда; C – ложный ноль, то есть вариант, имеющий либо наибольшую частоту, либо равноудалённый от максимального и минимального вариантов. Тогда основные числовые характеристики для вариантов x и u связаны соотношениями:

$$\overline{x} = h \cdot \overline{u} + C;$$

$$S^2(x) = h^2 \cdot S^2(x);$$

$$\tilde{A}(x) = \tilde{A}(u);$$

$$\tilde{E}(x) = \tilde{E}(u).$$

Пример 91. Перейдя к условным вариантам, вычислить эмпирические коэффициенты асимметрии и эксцесса для интервального вариационного ряда

Интервалы	[3–7)	[7–11)	[11–15)	[15–19)	[19–23)	[23–27)
Частоты	6	9	11	12	8	4

Решение. Шаг интервального вариационного ряда равен $h = 4$. Интервал, имеющий наибольшую частоту, – это интервал $[15;\ 19]$, тогда условный нуль это его середина, то есть $C = 17$. Перейдём к условным вариантам по формуле $u_i = \dfrac{x_i - 17}{4}$. Тогда получим дискретный вариационный ряд:

u_i	–3	–2	–1	0	1	2
m_i	6	9	11	12	8	4

Найдём эмпирические начальные моменты:

$$\tilde{V}_1 = \frac{1}{n}\sum_{i=1}^{n} u_i \cdot m_i =$$

$$= \frac{1}{50}\left((-3)\cdot 6 + (-2)\cdot 9 + (-1)\cdot 11 + 0 \cdot 12 + 1 \cdot 8 + 2 \cdot 4\right) = -\frac{31}{50} = -0{,}62;$$

$$\tilde{V}_2 = \frac{1}{n}\sum_{i=1}^{n} u_i^2 \cdot m_i =$$

$$= \frac{1}{50}\left((-3)^2 \cdot 6 + (-2)^2 \cdot 9 + (-1)^2 \cdot 11 + 0^2 \cdot 12 + 1^2 \cdot 8 + 2^2 \cdot 4\right) =$$

$$= \frac{125}{50} = 2{,}5;$$

$$\tilde{V}_3 = \frac{1}{n}\sum_{i=1}^{n} u_i^3 \cdot m_i =$$

$$= \frac{1}{50}\left((-3)^3 \cdot 6 + (-2)^3 \cdot 9 + (-1)^3 \cdot 11 + 0^3 \cdot 12 + 1^3 \cdot 8 + 2^3 \cdot 4\right) =$$

$$= -\frac{205}{50} = -4{,}1;$$

$$\tilde{V}_4 = \frac{1}{n}\sum_{i=1}^{n} u_i^4 \cdot m_i =$$

$$= \frac{1}{50}\left((-3)^4 \cdot 6 + (-2)^4 \cdot 9 + (-1)^4 \cdot 11 + 0^4 \cdot 12 + 1^4 \cdot 8 + 2^4 \cdot 4\right) =$$

$$= \frac{713}{50} = 14{,}26.$$

Найдём эмпирические центральные моменты:

$$\tilde{\mu}_2 = \tilde{V}_2 - \tilde{V}_1^2 = 2{,}5 - (-0{,}62)^2 = 2{,}11156;$$

$$\tilde{\mu}_3 = \tilde{V}_3 - 3\tilde{V}_2 \cdot \tilde{V}_1 + \tilde{V}_1^3 = -4{,}1 - 3 \cdot 2{,}5 \cdot (-0{,}62) + (-0.62)^3 \approx -1{,}1554;$$

$$\tilde{\mu}_4 = \tilde{V}_4 - 4\tilde{V}_3 \cdot \tilde{V}_1 + 6\tilde{V}_2 \cdot \tilde{V}_1^2 - 3\tilde{V}_1^4 =$$

$$= 14{,}36 - 4 \cdot (-4{,}1) \cdot (-0{,}62) + 6 \cdot 2{,}5 \cdot (-0{,}62)^2 -$$

$$-3(-0{,}62)^4 \approx 7{,}1956.$$

Тогда среднее арифметическое равно

$$\bar{x} = h \cdot \bar{u} + C = h \cdot \tilde{V}_1 + C = 4 \cdot (-0{,}62) + 17 = 14{,}52.$$

Эмпирическая дисперсия равна

$$S^2(x) = h^2 \cdot S^2(u) = h^2 \cdot \tilde{\mu}_2 = 16 \cdot 2{,}1156 = 33{,}8496.$$

Эмпирический коэффициент асимметрии равен

$$\tilde{A} = \frac{\tilde{\mu}_3}{\sqrt{\tilde{\mu}_2^3}} = \frac{-1{,}1554}{\sqrt{2{,}1156^3}} \approx -0{,}375.$$

Эмпирический коэффициент эксцесса равен

$$\tilde{E} = \frac{\tilde{\mu}_4}{\tilde{\mu}_2^2} - 3 = \frac{7{,}1956}{2{,}1156^2} - 3 \approx -1{,}392.$$

2.2.2. Статистическое оценивание параметров распределения

В самом общем смысле содержание этой темы можно сформулировать как совокупность методов, позволяющих делать научно обоснованные выводы о числовых параметрах распределения генеральной совокупности по случайной выборке из неё. Если, например, нас интересует математическое ожидание генеральной совокупности, то задача статистической оценки параметров заключается в том, чтобы найти такую выборочную характеристику, которая позволила бы получить по возможности более точное и надёжное представление об интересующем нас параметре (в данном случае о математическом ожидании). Состав выборки случаен, поэтому выводы о параметрах генеральной совокупности, сделанные по выборочным данным, могут быть ложными. С возрастанием числа элементов выборки вероятность правильного вывода увеличивается. Поэтому всякому решению, принимаемому при статистической оценке параметров, стараются поставить в соответствие вероятность, характеризующую степень достоверности принимаемого решения.

Сформулируем задачу оценки параметров в общем виде. Пусть X – случайная величина, подчинённая закону распределения $F(X;\Theta)$, где Θ – параметр распределения, числовое значение которого неизвестно. Исследовать все элементы генеральной совокупности для вычисления параметра Θ не представляется возможным, поэтому об этом параметре пытаются судить по выборкам из генеральной совокупности.

Всякую однозначно определённую функцию результатов наблюдений, с помощью которой судят о значении параметра Θ, называют *оценкой* (или *статистикой*) параметра Θ. Рассмотрим некоторое множество выборок, объёмом n каждая. Выборочную оценку параметра Θ, вычисленную по i-й выборке, обозначим Θ_{n_i}. Так как состав выборки случаен, то можно сказать, что Θ_{n_i} примет неизвестное заранее числовое значение, то есть является случайной величиной.

Известно, что случайная величина определяется законом распределения и числовыми характеристиками, следовательно, и выборочную оценку можно также описывать законом распределения и числовыми характеристиками.

Для того чтобы отразить случайный характер выборки объёмом n из генеральной совокупности, обозначим её $X_1, X_2, ..., X_n$, а выборочную оценку параметра Θ – через $\tilde{\Theta}_n$. Следовательно, можно записать $\tilde{\Theta}_n = f(X_1, X_2, ..., X_n)$. Выбор оценки, позволяющей получить хорошее приближение оцениваемого параметра, – основная задача теории оценивания.

2.2.3. Основные свойства оценок

Оценку $\tilde{\Theta}_n$ параметра Θ называют *несмещённой*, если её математическое ожидание равно оцениваемому параметру Θ, то есть $M(\tilde{\Theta}_n) = \Theta$. Если это равенство не выполняется, то оценка $\tilde{\Theta}_n$ может либо завышать значение Θ, либо занижать его. В обоих случаях это приводит к систематическим ошибкам в оценке параметра. Требование несмещённости гарантирует отсутствие систематических ошибок при оценке параметров.

Несмещённую оценку $\tilde{\Theta}_n$, которая имеет наименьшую дисперсию среди всех несмещённых оценок параметра Θ, вычисленных по выборкам одного и того же объёма, называют *эффективной оценкой*.

Оценку $\tilde{\Theta}_n$ параметра Θ называют *состоятельной*, если она подчиняется закону больших чисел, то есть при достаточно большом числе независимых наблюдений n с вероятностью, близкой к единице, можно утверждать, что разность между $\tilde{\Theta}_n$ и Θ по абсолютной величине окажется меньше любого малого числа τ, или $p\left(\left|\tilde{\Theta}_n - \Theta\right| < \tau\right) > 1 - \varepsilon$, где ε – положительное число, близкое нулю.

На практике при оценке параметров не всегда удаётся удовлетворить одновременно требованиям не смещённости, эффективности и состоятельности оценки. Так, например, может оказаться, что для простоты расчётов целесообразно использовать незначительно смещённую оценку. Однако выбору оценки всегда должно предшествовать её критическое рассмотрение со всех точек зрения.

2.2.4. Оценка математического ожидания и дисперсии

Наиболее важными числовыми характеристиками случайной величины являются математическое ожидание и дисперсия. Рассмотрим вопрос о том, какие выборочные характеристики лучше всего в смысле не смещённости, эффективности и состоятельности оценивают математическое ожидание и дисперсию.

Теорема 9. Средняя арифметическая $\bar{x}$, вычисленная по n независимым наблюдениям над случайной величиной X, которая имеет математическое ожидание μ, является несмещённой оценкой этого параметра.

Теорема 10. Средняя арифметическая $\bar{x}$, вычисленная по n независимым наблюдениям над случайной величиной, которая имеет математическое ожидание μ и дисперсию σ^2, является состоятельной оценкой этого параметра.

На практике часто используют следующее утверждение: если случайная величина X распределена по нормальному закону с параметрами $\left(\mu; \sigma^2\right)$, то несмещённая оценка $\overline{x}$ математического ожидания μ имеет минимальную дисперсию $\dfrac{\sigma^2}{n}$, поэтому средняя арифметическая $\overline{x}$ в этом случае является эффективной оценкой математического ожидания μ.

Теорема 11. Если случайная выборка состоит из n независимых наблюдений над случайной величиной X с математическим ожиданием μ и дисперсией σ^2, то выборочная дисперсия $S^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ не является несмещённой оценкой генеральной дисперсии.

Несмещённой оценкой дисперсии генеральной совокупности является $\hat{S}^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)^2$. Обычно эту оценку называют *исправленной выборочной дисперсией*. Дробь $\dfrac{n}{n-1}$ называют поправкой Бесселя. Тогда имеем равенство $\hat{S}^2 = \dfrac{n}{n-1}S^2$. При малых значениях n поправка Бесселя довольно сильно отличается от единицы, с увеличением n она быстро стремится к единице. При $n > 50$ практически нет разницы между S^2 и $\hat{S}^2$. Можно показать, что оценки S^2 и $\hat{S}^2$ являются состоятельными оценками σ^2.

Несмещённой, состоятельной и эффективной оценкой σ^2 является оценка $\hat{S}_x^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}\left(x_i - \mu\right)^2$, для вычисления которой необходимо знать математическое ожидание. Заметим, что оценка $\hat{S}_x^2$ эффективна лишь при условии нормальности распределения случайной величины X в генеральной совокупности. Оценки S^2 и $\hat{S}^2$не являются эффективными. В том случае, когда значение математического ожидания неизвестно, для оценки дисперсии σ^2 используют состоятельную и несмещённую оценку $\hat{S}^2$.

Пример 92. Из генеральной совокупности получена некоторая выборка обьёма $n = 100$:

x_i	50	80	90	110	120	140	170
n_i	12	7	6	6	15	24	30

Найти несмещённую оценку генерального среднего.

Решение. Несмещённую оценкой генерального среднего $\overline{x_Г}$ является выборочное среднее $\overline{x_В}$. Поскольку варианты выборки являются большими числами, перейдём к условным вариантам

$$u_i = \frac{x_i - 110}{10}$$

(в качестве C выбрано значение четвёртой варианты $x_4 = 110,$ а в качестве b-число 10, поскольку все варианты кратны десяти):

$$u_1 = \frac{x_1 - 110}{10} = \frac{50 - 110}{10} = -6; \quad u_2 = -3; \quad u_3 = -2; \quad u_4 = 0; \quad u_5 = 1; \quad u_6 = 3; \quad u_7 = 6;$$

Теперь легко найти выборочное среднее:

$$\overline{x}_n = b \cdot \frac{\sum_{i=1}^{7} n_i u_i}{n} + C = 10 \cdot \frac{\sum_{i=1}^{7} n_i u_i}{100} + 110 = \frac{162}{10} + 110 = 126{,}2.$$

Пример 93. Из генеральной совокупности получена некоторая выборки объема $n = 100$.

x_i	0,002	0,004	0,006	0,008	0,01	0,012	0,014
n_i	7	29	35	12	9	5	3

Найти смещённую оценку генеральной дисперсии.

Решение. Смещённую оценку генеральной дисперсии $D_Г$ является выборочная дисперсия $D_В$. Поскольку варианты выборки являются малыми близкими к нулю числами, перейдём к условным вариантам

$$u_i = x_i \cdot 1000$$

(в качестве b выбрано число 1000, поскольку в таком случае мы получим целые числа):

$$u_1 = x_1 \cdot 1000 = 0{,}002 \cdot 1000 = 2; \quad u_2 = x_2 \cdot 1000 = 0{,}004 \cdot 1000 = 4;$$

$$u_3 = x_3 \cdot 1000 = 0{,}006 \cdot 1000 = 6; \quad u_4 = x_4 \cdot 1000 = 0{,}008 \cdot 1000 = 8;$$

$$u_5 = x_5 \cdot 1000 = 0{,}01 \cdot 1000 = 10; \quad u_6 = x_6 \cdot 1000 = 0{,}012 \cdot 1000 = 12;$$

$$u_7 = x_7 \cdot 1000 = 0{,}014 \cdot 1000 = 14;$$

Теперь легко найти выборочную дисперсию:

$$D_n = \frac{\sum_{i=1}^{7} n_i u_i^2}{b^2 n} - \frac{(\sum_{i=1}^{7} n_i u_i)^2}{(bn)^2} = 5{,}5216 \cdot 10^{-6}.$$

Пример 94. По данным выборки обьёма $n = 21$ найдена выборочная дисперсия $D_n = 5$

Найти несмещённую оценку генеральной дисперсии.

Решение. Несмещённую оценку генеральной дисперсии является исправленная дисперсия:

$$s^2 = \frac{n}{n-1} D_n = \frac{21}{21-1} \cdot 5 = 5{,}25.$$

2.2.5. Метод максимального правдоподобия

Основным способом получения оценок параметров генеральной совокупности по данным выборке является метод максимального правдоподобия. Основная идея метода заключается в следующем.

Пусть $X_1, X_2, ..., X_n$ – результаты независимых наблюдений над случайной величиной X, которая может быть как дискретной, так и непрерывной; $f(X;\Theta)$ – вероятность значения (если случайная величина дискретна) и плотность вероятности (если случайная величина непрерывна). Функция $f(X;\Theta)$ зависит от неизвестного параметра Θ, который требуется оценить по выборке.

Если $X_1, X_2, ..., X_n$ – независимые случайные величины, то функцией правдоподобия называется выражение $L = f(X_1;\Theta) \cdot f(X_2;\Theta) \cdot ... \cdot f(X_n;\Theta)$. В качестве оценки неизвестного параметра Θ берётся такое значение $\widetilde{\Theta}$, при подстановке которого вместо параметра Θ получаем максимальное значение функции L. Оценку $\tilde{\Theta}$ обычно называют *оценкой максимального правдоподобия*. Оценка $\tilde{\Theta}$ зависит от количества и числовых значений случайных величин X_i, следовательно, сама является случайной величиной.

При максимизации функции L подразумевается, что значения $X_1, X_2, ..., X_n$ фиксированы, а переменной является параметр Θ (иными словами, максимум отыскивается в предположении, что X_i заменены их числовыми значениями). Если L дифференцируема относительно параметра Θ то для отыскания максимума надо решить уравнение $\frac{\partial L}{\partial \Theta} = 0$. В качестве оценки Θ выбрать решение, которое обращает функцию L в максимум. Иногда удобно рассматривать уравнение $\frac{\partial \ln L}{\partial \Theta} = 0$.

Согласно методу максимального правдоподобия для нормально распределённой генеральной совокупности в качестве оценок математического ожидания и дисперсии нужно брать соответственно среднюю арифметическую и эмпирическую дисперсию S^2.

Пример 95. Найти методом максимального правдоподобия по выборке $X_1, X_2, ..., X_n$ точечную оценку неизвестного параметра λ показательного распределения, плотность которого $f(x) = \lambda \cdot e^{-\lambda x}, \quad (x \geq 0)$.

Решение. Составим функцию правдоподобия $L = \left(\lambda \cdot e^{-\lambda x_1}\right) \cdot \left(\lambda \cdot e^{-\lambda x_2}\right) \cdot \ldots \cdot \left(\lambda \cdot e^{-\lambda x_n}\right) = \lambda^n \cdot e^{-\lambda \sum_{i=1}^{n} x_i}$. Прологарифмируем функцию правдоподобия $\ln L = n \cdot \ln \lambda - \lambda \sum_{i=1}^{n} x_i$. Найдём первую производную по λ: $\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$. Отсюда $\lambda = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$. Так как $\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{n}{\lambda^2} = -n \cdot \bar{x}^{-2} < 0$ в силу положительности $\bar{x}$, то оценкой метода максимального правдоподобия параметра λ является величина, обратная среднему арифметическому.

Пример 96. Случайная величина X (количество битой стеклянной посуды в одной упаковки) распределена по закону Пуассона с неизвестным параметром λ. В результате статистических исследований получено такое эмпирическое распределение количества битой стеклянное посуды в $n = 1000$ упаковках:

x_i	0	1	2	3	4	5	6
n_i	554	324	98	19	3	1	1

Найти методом наибольшего правдоподобия точечную оценку неизвестного параметра λ распределения Пуассона.

Решение. Поскольку случайная величина X распределена по закону Пуассона, то функция правдоподобия имеет вид

$$L = \left(\frac{\lambda^0 e^{-\lambda}}{0!}\right)^{554}\left(\frac{\lambda^1 e^{-\lambda}}{1!}\right)^{324}\left(\frac{\lambda^2 e^{-\lambda}}{2!}\right)^{98}\left(\frac{\lambda^3 e^{-\lambda}}{3!}\right)^{19}\left(\frac{\lambda^4 e^{-\lambda}}{4!}\right)^{3}\left(\frac{\lambda^5 e^{-\lambda}}{5!}\right)^{1}\left(\frac{\lambda^6 e^{-\lambda}}{6!}\right)^{1} = \frac{\lambda^{600} e^{-1000\lambda}}{2^{133}\cdot 3^{25}\cdot 5^2}.$$

Запишем логарифмическую функцию правдоподобия:

$$\ln L = 600\ln\lambda - 1000\lambda - \ln(2^{133}\cdot 3^{25}\cdot 5^2).$$

Определим производную логарифмической функции распределения по λ:

$$\frac{d\ln L}{d\lambda} = \frac{600}{\lambda} - 1000.$$

Приравняв её к нулю, найдём единственный корень уравнения правдоподобия:

$$\lambda^* = 0{,}6.$$

Поскольку вторая производная логарифмической функции правдоподобия

$$\frac{d^2\ln L}{d\lambda^2} = -\frac{600}{\lambda^2} < 0$$

всегда отрицательная, точечной максимального правдоподобия параметра λ распределения Пуассона будет

$$\lambda^* = 0{,}6.$$

Пример 97. Случайная величина X (длина детали) распределена по нормальному закону с неизвестными параметрами α, σ. В результате статистических исследований получено такое эмпирическое распределение длины $n=1000$ деталей:

Длина детали,см	5,1	5,2	5,3	5,4	5,5	5,6	5,7
Количество деталей	7	78	289	392	195	36	3

Найти методом наибольшего правдоподобия точечную оценку неизвестных парамтров α, σ нормального распределения.

Решение. Плотность распределения вероятностей случайной величины X

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

Запишем фу нкцию правдоподобия:

$$L=\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5.1-a)^2}{2\sigma^2}}\right)^7\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5.2-a)^2}{2\sigma^2}}\right)^{78}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5.3-a)^2}{2\sigma^2}}\right)^{289}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5.4-a)^2}{2\sigma^2}}\right)^{392}\times$$

$$\times\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5.5-a)^2}{2\sigma^2}}\right)^{195}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5.6-a)^2}{2\sigma^2}}\right)^{36}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5.7-a)^2}{2\sigma^2}}\right)^{3}=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{1000}\times$$

$$\times\exp\{[-7\cdot(5.1-a)^2+78\cdot(5.2-a)^2+289\cdot(5.3-a)^2+392\cdot(5.4-a)^2+195\cdot(5.5-a)^2+36\cdot(5.6-a)^2+$$

$$+3\cdot(5.7-a)^2]/[2\sigma^2]\}=(2\pi\sigma^2)^{-500}\cdot e^{-\frac{1000a^2-10762a+28965.1}{2\sigma^2}}.$$

Запишем логарифмическую функцию правдоподобия

$$\ln L=-500\cdot\ln 2\pi\sigma^2-\frac{1000a^2-10762a+28965.1}{2\sigma^2}$$

Определим частные производные логарифмической функции распределения по α и σ^2:

$$\frac{d\ln L}{da}=-\frac{2000a-10762}{2\sigma^2};$$

$$\frac{d\ln L}{d\sigma^2}=-\frac{500}{\sigma^2}+\frac{1000a^2-10762a+28965.1}{2\sigma^2}.$$

Приравняв их к нулю, найдём решение системы уравнений

$$\begin{cases} -\dfrac{2000a-10762}{2\sigma^2}=0 \\ -\dfrac{500}{\sigma^2}+\dfrac{1000a^2-1076a+28965.1}{2\sigma^2}=0 \end{cases}$$

Решением будет: a=5,381; $\sigma^2=0,009939\approx 0,1.$

Итак, точечными оценками максимального правдоподобия параметров a, σ^2 нормального распределения являются

$$a^*=5,381;\quad \sigma^*=\sqrt{0,009939}\approx 0,1.$$

2.2.6. Метод наименьших квадратов

Изложенный ранее метод максимального правдоподобия всегда приводит к состоятельным оценкам, хотя иногда смещённым. Известно, что этот метод использует наилучшим образом всю информацию о неизвестном параметре, содержащуюся в выборке. Однако часто его применение связано с необходимостью решения сложных систем уравнений.

Другим методом, имеющим большое практическое применение в задачах оценивания неизвестных параметров генеральной совокупности по выборке и часто приводящим к более простым выкладкам, является *метод наименьших квадратов*.

Пусть, как и прежде, X – случайная величина (дискретная или непрерывная) с законом распределения $f(X;\Theta)$, где Θ – неизвестный параметр генеральной совокупности, который нужно оценить по выборке; $X_1, X_2, ..., X_n$ – независимые наблюдения; $\tilde{\Theta}$ – оценка параметра Θ, зависящая от количества наблюдений и их числовых значений.

Основная идея метода наименьших квадратов в приложении к оцениванию параметров сводится к тому, чтобы в качестве оценки неизвестного параметра принимать значение, которое минимизирует сумму квадратов отклонений между оценкой и параметром для всех наблюдений. То есть находится минимум функции $F=\sum_{i=1}^{n}\left(\Theta-\hat{\Theta}(x_i)\right)^2$.

Если исходная случайная величина имеет нормальный закон распределения, то метод максимального правдоподобия и метод наименьших квадратов дают одинаковые результаты.

Особенно часто метод наименьших квадратов применяется в задачах выравнивания или сглаживания. Пусть в результате наблюдений получен ряд точек с координатами $(x_1;\, y_1), (x_2;\, y_2), \ldots, (x_n;\, y_n)$. Если заранее известно, что зависимость между переменными имеет вид $y = f(x;\, a_1;\, a_2)$, то необходимо определить числовые параметры a_1, a_2, которые наилучшим образом, в смысле наименьших квадратов, описывали бы зависимость, полученную при наблюдении. То есть найти минимум функции $F = \sum_{i=1}^{n}\left(y_i - f(x;\, a_1;\, a_2)\right)^2$. Для этого нужно решить систему уравнений

$$\begin{cases} \dfrac{\partial F}{\partial a_1} = 0; \\ \dfrac{\partial F}{\partial a_2} = 0. \end{cases}$$

Пример 98. Найти методом наименьших квадратов коэффициенты линейной зависимости $y = a + bx$ по полученным эмпирическим точкам с координатами $(x_1;\, y_1), (x_2;\, y_2), \ldots, (x_n;\, y_n)$.

Решение. Функция F имеет вид $F = \sum_{i=1}^{n}\left(y_i - (a + bx)\right)^2$. Запишем систему уравнений для нахождения неизвестных параметров

$$\begin{cases} \dfrac{\partial F}{\partial a} = -2\sum\limits_{i=1}^{n}\left(y_i - a - bx_i\right) = 0; \\ \dfrac{\partial F}{\partial b} = -2\sum\limits_{i=1}^{n}\left(y_i - a - bx_i\right)\cdot x_i = 0; \end{cases}$$

или

$$\begin{cases} \sum\limits_{i=1}^{n} y_i - a\cdot n - b\sum\limits_{i=1}^{n} x_i = 0; \\ \sum\limits_{i=1}^{n} y_i\cdot x_i - a\sum\limits_{i=1}^{n} x_i - b\sum\limits_{i=1}^{n} x_i^n = 0. \end{cases}$$

Выражая из первого уравнения системы параметр a, получим

$$a=\frac{1}{n}\left(\sum_{i=1}^{n}y_i-b\sum_{i=1}^{n}x_i\right),$$

подставим во второе уравнение системы

$$n\sum_{i=1}^{n}y_ix_i-bn\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i+b\left(\sum_{i=1}^{n}x_i\right)^2=0,$$

отсюда

$$b=\frac{n\sum\limits_{i=1}^{n}x_iy_i-\sum\limits_{i=1}^{n}x_i\sum\limits_{i=1}^{n}y_i}{n\sum\limits_{i=1}^{n}x_i^2-\left(\sum\limits_{i=1}^{n}x_i\right)^2},$$

подставляя полученное в выражение для a, находим его. Используя понятие средней арифметической, результат можно записать гораздо компактней:

$$b=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\overline{x^2}-\left(\overline{x}\right)^2};\qquad a=\overline{y}-b\cdot\overline{x}.$$

Пример 99. Методом наименьших квадратов для данных, представленных в таблице, найти линейную зависимость $y=ax+b$.

Данные

i	1	2	3	4	5	6	7	8
x_i	-2	-1	0	1	2	3	4	5
y_i	-11,47	-7,59	-4,32	-0,41	3,01	6,91	10,12	14,08

Решение. Параметры a и b по методу наименьших квадратов можно найти из системы уравнений.

$$\begin{cases} a\sum x_i^2+b\sum x_i=\sum x_iy_i \\ a\sum x_i+bn=\sum y_i \end{cases}$$

где суммирование ведется по i от 1 до *n*, *n*=8. Составим расчетную таблицу:

i	1	2	3	4	5	6	7	8	сумма
x_i	-2	-1	0	1	2	3	4	5	12

y_i	-11,47	-7,59	-4,32	-0,41	3,01	6,91	10,02	14,08	10,33
x_i^2	4	1	0	1	4	9	16	25	60
$x_i\ y_i$	22,94	7,59	0	-0,41	6,02	20,73	40,48	70,4	167,75

Получаем систему:

$$\begin{cases} 60a + 12b = 167{,}75 \\ 12a + 8b = 10{,}33 \end{cases}$$

Откуда находим a=3,625, b=-4.146, то есть получаем функцию $y = 3{,}625x - 4{,}146$.

Пример 100. Экспериментальные данные о значениях переменных x и y приведены в таблице:

χ_i	1	2	4	6	8
y_i	3	2	1	0.5	0

В результате их выравнивания получена функция $y = \frac{5}{2x}$. Используя метод наименьших квадратов, аппроксимировать эти данные линейной зависимостью $y = ax + b$ (найти параметры a и b).Выяснить, какая из двух линий лучше (в смысле метода наименьших квадратов) выравнивает экспериментальные данные. Сделать чертеж.

Решение. Параметры a и b по методу наименьших квадратов можно найти из системы уравнений:

$$\begin{cases} a\sum x_i^2 + b\sum x_i = \sum x_i y_i \\ a\sum x_i + bn = \sum y_i \end{cases}$$

где суммирование ведется по i от 1 до n,n=8. Составим расчетную таблицу:

i	1	2	3	4	5	6	7	8	сумма
x_i	-2	-1	0	1	2	3	4	5	12
y_i	-11,47	-7,59	-4,32	-0,41	3,01	6,91	10,02	14,08	10,33
x_i^2	4	1	0	1	4	9	16	25	60
$x_i\ y_i$	22,94	7,59	0	-0,41	6,02	20,73	40,48	70,4	167,75

Получаем систему:

$$\begin{cases} 121a + 21b = 14 \\ 21a + 5b = 6.5 \end{cases}$$

Откуда находим a=-0,405, b=3.003, то есть получаем функцию $y = -0{,}405x + 3{,}003$.

Выясним, какая из двух линий лучше (в смысле метода наименщих квадратов) выравнивает экспериментальние данные.

Обозначим $y_1 = \dfrac{5}{2x}$, $y_2 = -0{,}405x + 3{,}003$. Вычислим сумму квадратов отклонений в обоих случаях:

Сумма

x_i	1	2	4	6	8	21
y_i	3	2	1	0,5	0	6,5
y_{1i}	2,500	1,250	0,625	0,417	0,313	5,104
y_{2i}	2,598	2,193	1,383	0,573	-0,237	6,510
$(y_i - y_{1i})^2$	0,250	0,563	0,141	0,007	0,098	1,058
$(y_i - y_{2i})^2$	0,162	0,037	0,147	0,005	0,056	0,407

Видно, что так 0,407<1,058, вторая линия лучше в смысле методы наименьших квадратов выравнивает данные (Рис. 6).

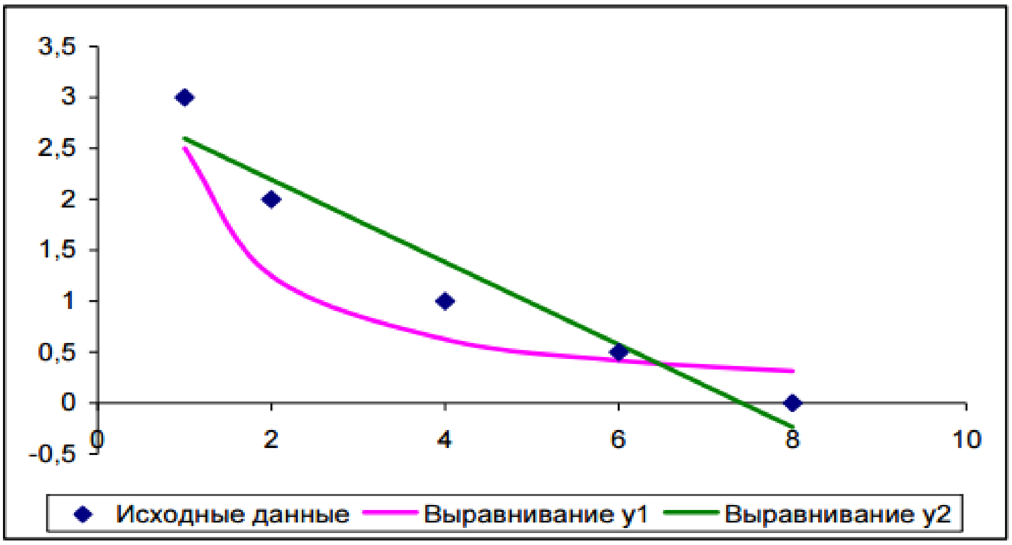

Рис. 6. Сравнение выравнивание.

2.2.7. Распределение средней арифметической для выборок из нормальной совокупности

Распределение Стьюдента

Выборочная средняя, вычисленная по конкретной выборке, есть определённое число. Так как состав выборки случаен, то средняя арифметическая, вычисленная для элементов другой выборки того же объёма из той же генеральной совокупности, определяется числом, как правило, отличным от первого, то есть средняя меняется от выборки к выборке.

Следовательно, выборочную среднюю можно рассматривать как случайную величину, что позволяет говорить о законе распределения выборочной средней.

Теорема 12. Если случайная величина X подчиняется нормальному закону распределения с параметрами $\left(\mu; \sigma^2\right)$, а $X_1, X_2, ..., X_n$ – ряд независимых наблюдений над случайной величиной X, каждое из которых имеет те же характеристики, что X, то выборочная средняя $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ также подчиняется нормальному закону распределения с параметрами $\left(\mu; \frac{\sigma^2}{n}\right)$.

Нормированное отклонение $\frac{\left(\bar{x}-\mu\right)\sqrt{n}}{\sigma}$ подчиняется нормальному закону распределения со средним значением, равным нулю, и дисперсией, равной единице. Действительно, используя свойства математического ожидания, а также тот факт, что $\bar{x}$ и μ независимы, имеем:

$$M\left(\frac{\left(\bar{x}-\mu\right)\cdot\sqrt{n}}{\sigma}\right) = \frac{\sqrt{n}}{\sigma}\left(M\left(\bar{x}\right)-\mu\right) = \frac{\sqrt{n}}{\sigma}\left(\mu-\mu\right) = 0$$

и

$$D\left(\frac{\left(\bar{x}-\mu\right)\cdot\sqrt{n}}{\sigma}\right) = \frac{n}{\sigma^2}\left(D\left(\bar{x}\right)-D(\mu)\right) = \frac{n}{\sigma^2}\cdot\frac{\sigma^2}{n} = 1.$$

Пример 101. Автомат штампует детали. Контролируется длина детали, которая подчиняется нормальному закону распределения. Найти вероятность того, что средняя длина $\overline{x}$ деталей, отобранных случайным образом, отклонится от математического ожидания более чем на 2 мм, если дисперсия случайной величины X равна $\sigma^2 = 9$ мм^2, а количество деталей в выборке $n = 16$.

Решение. Случайная величина X имеет нормальное распределение с математическим ожиданием μ и дисперсией $\sigma_{\overline{x}}^2 = \frac{\sigma^2}{n} = \frac{9}{16}$, то есть $\sigma_{\overline{x}} = \frac{3}{4}$. Найдём вероятность того, что при $\left|\overline{x} - \mu\right| < 2$ она равна $p\left(\left|\overline{x} - \mu\right| < 2\right) = 2\Phi\left(\frac{2}{\sigma_x}\right) = 2\Phi\left(\frac{8}{3}\right) \approx 0{,}9982$, следовательно:

$$p\left(\left|\overline{x} - \mu\right| \geq 2\right) \approx 1 - 0{,}9982,$$

то есть практически можно быть уверенным, что наблюдаемая средняя длина детали отклонится от заданной не более чем на 2 мм.

Итак, если случайная величина X имеет нормальное распределение, то нормированное отклонение $\frac{\left(\overline{x} - \mu\right)\sqrt{n}}{\sigma}$ также подчиняется нормальному закону распределения. Однако дисперсия генеральной совокупности σ^2 почти всегда оказывается неизвестной, поэтому вызывает большой практический интерес изучение распределения статистики $t = \frac{\left(\overline{x} - \mu\right)\sqrt{n}}{\hat{S}^2}$, где $\hat{S}^2$ – несмещённая и состоятельная оценка дисперсии, вычисленная по выборочным данным. Распределение статистики t не зависит ни от математического ожидания μ случайной величины X, ни от дисперсии σ^2, а лишь зависит от объёма выборки n. Закон распределения статистики t называют *распределением Стьюдента*. Распределение Стьюдента табулировано во всех учебниках по математической статистике.

Из анализа распределения Стьюдента при $n > 50$ видно, что оно мало отличается от нормального.

2.2.8. Распределение дисперсии в выборках из нормальной генеральной совокупности

Распределение χ^2 *Пирсона.*

Рассмотрим закон распределения выборочной дисперсии, рассчитанной для наблюдений, взятых из нормальной генеральной совокупности. Так как состав выборки подвержен случайности, то выборочную дисперсию, как и $\bar{x}$, следует рассматривать как случайную величину и говорить о законе распределения выборочной дисперсии. При анализе распределения выборки следует иметь в виду два случая:

1) математическое ожидание случайной величины известно;

2) математическое ожидание неизвестно.

Случай 1. Предположим, что математическое ожидание случайной величины известно. Условимся считать, что случайная величина X подчиняется нормальному закону распределения с параметрами $(\mu; \sigma^2)$, а $X_1, X_2, ..., X_n$ – ряд независимых наблюдений, каждое из которых подчиняется нормальному закону распределения с математическим ожиданием μ и дисперсией σ^2. Тогда выборочная дисперсия вычисляется по формуле

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 .$$

Разделим обе части этого равенства на σ^2 и умножим на n. Получим $\frac{n \cdot S_x^2}{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2 = \sum_{i=1}^{n} t_i^2$. Статистика t имеет нормальный закон распределения с параметрами $M(t) = 0$ и $D(t) = 1$.

Пусть $\chi^2 = \frac{n \cdot S_s^2}{\sigma^2} = \sum_{i=1}^{n} t_i^2$.

Случайная величина, представляющая собой сумму квадратов независимых случайных величин, каждая из которых подчиняется нормальному закону.

распределения с параметрами $(\mu = 0;\ \sigma^2 = 1)$, называется *случайной величиной с распределением* χ^2 и $k = n$ степенями свободы.

Распределение статистики χ^2 не зависит ни от математического ожидания μ случайной величины X, ни от дисперсии σ^2, а зависит лишь от объёма выборки n. Найдём математическое ожидание распределения χ^2:

$$M\left(\chi^2\right) = M\left(\sum_{i=1}^{n} t_i^2\right) = \sum_{i=1}^{n} M\left(t_i^2\right) = \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2 =$$

$$= \frac{1}{\sigma^2} \cdot \frac{n}{n} \sum_{i=1}^{n}\left(x_i - \mu\right)^2 = \frac{n}{\sigma^2} \cdot \sigma^2 = n.$$

Следовательно, математическое ожидание случайной величины с распределением χ^2 и $k = n$ степенями свободы равно числу степеней свободы. В специальной литературе можно найти доказательство того, что дисперсия распределения χ^2 равна удвоенному числу степеней свободы. Дифференциальная функция распределения χ^2 сложна, и интегрирование её является весьма трудоёмким процессом, поэтому составлены таблицы распределения χ^2.

Случай 2. Рассмотрим закон распределения выборочной дисперсии, когда математическое ожидание случайной величины неизвестно. Как и прежде, случайная величина подчиняется нормальному закону распределения с параметрами $(\mu; \sigma^2)$, а $X_1, X_2, \ldots, X_n$ – ряд независимых наблюдений, каждое из которых подчиняется нормальному закону распределения с математическим ожиданием μ и дисперсией σ^2. Тогда исправленная дисперсия выборки вычисляется по формуле $\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$. Примем без доказательства тот факт, что случайная величина $\frac{n \cdot \hat{S}^2}{\sigma^2}$ имеет распределение χ^2 с $k = n-1$ степенями свободы.

2.2.9. Понятие доверительного интервала

Доверительная вероятность

Оценку неизвестного параметра генеральной совокупности одним числом называют *точечной оценкой*. Наряду с точечным оцениванием статистическая теория занимается вопросами интервального оценивания. Задачу интервального оценивания в самом общем случае можно сформулировать так: по данным выборки построить числовой интервал, относительно которого с заранее выбранной точностью можно сказать, что внутри находится числовой параметр. *Интервальное оценивание* особенно необходимо при малом числе наблюдений, когда точечная оценка мало надёжна.

Доверительным интервалом $\left[\theta_{1;\,n};\,\theta_{2;\,n}\right]$ для параметра θ называют такой интервал, относительно которого можно с заранее выбранной вероятностью $p = 1-\alpha$, близкой к единице, утверждать, что содержит неизвестное значение параметра θ, то есть $p\left(\theta_{1;\,n} < \theta < \theta_{2;\,n}\right) = 1-\alpha$.

Чем меньше для выбранной вероятности разность $\left(\theta_{2;\,n} - \theta_{1;\,n}\right)$, тем точнее оценка неизвестного параметра θ, и наоборот, если этот интервал велик, то оценка, произведённая с его помощью, мало пригодна для практики. Концы доверительного интервала $\theta_{1;n}$ и $\theta_{2;n}$ зависят от элементов выборки, поэтому их значения могут меняться от выборки к выборке. Вероятность $p = 1-\alpha$ принято называть *доверительной вероятностью*, а число α – *уровнем значимости*.

2.2.10. Доверительный интервал для математического ожидания при известной дисперсии генеральной совокупности

Пусть случайная величина X распределена нормально, причём среднее квадратическое отклонение σ этого распределения известно. Требуется построить доверительный интервал для неизвестного математического ожидания μ с заданным уровнем значимости α. Ранее показано, что выборочное среднее распределено нормально с параметрами $M(\bar{x}) = \mu, \quad D(\bar{x}) = \dfrac{\sigma^2}{n}$, нормированное отклонение $\dfrac{(\bar{x}-\mu)\sqrt{n}}{\sigma}$ распределено также нормально с параметрами $M\left(\dfrac{(\bar{x}-\mu)\sqrt{n}}{\sigma}\right) = 0$ и $D\left(\dfrac{(\bar{x}-\mu)\sqrt{n}}{\sigma}\right) = 1$. Поэтому вероятность любого отклонения $|\bar{x}-\mu|$ может быть вычислена по формуле

$$p\left(\left|\frac{(\bar{x}-\mu)\sqrt{n}}{\sigma}\right| < z_{\text{кр}}\right) = 2\Phi(z_{\text{кр}}).$$

Для заданной доверительной вероятности имеем $p = 1-\alpha = 2\Phi(z_{\text{кр}})$ или $\Phi(z_{\text{кр}}) = \dfrac{1-\alpha}{2}$, далее по таблице функции $\Phi(x)$ находим $z_{\text{кр}}$. Преобразуем формулу к удобному виду

$$p\left(|\bar{x}-\mu| < \frac{z_{\text{кр}}\cdot\sigma}{\sqrt{n}}\right) = 2\Phi(z_{\text{кр}})$$

или

$$p\left(-\frac{z_{\text{кр}}\cdot\sigma}{\sqrt{n}} < \bar{x}-\mu < \frac{z_{\text{кр}}\cdot\sigma}{\sqrt{n}}\right) = 2\Phi(z_{\text{кр}}),$$

откуда

$$p\left(\bar{x}-\frac{z_{\text{кр}}\cdot\sigma}{\sqrt{n}} < \mu < \bar{x}+\frac{z_{\text{кр}}\cdot\sigma}{\sqrt{n}}\right) = 2\Phi(z_{\text{кр}}).$$

Таким образом, с вероятностью (надёжностью) $p = 1 - \alpha$ можно утверждать, что интервал $\left(\bar{x} - \frac{z_{\text{кр}} \cdot \sigma}{\sqrt{n}};\ \bar{x} + \frac{z_{\text{кр}} \cdot \sigma}{\sqrt{n}} \right)$ является доверительным для оценки математического ожидания.

Пример 102. Случайная величина X имеет нормальное распределение с известным средним квадратическим отклонением $\sigma = 12$. Найти доверительный интервал для оценки математического ожидания μ по выборочной средней $\bar{x} = 45$, если объём выборки $n = 36$ и доверительная вероятность $p = 0{,}95$.

Решение. Используя соотношение

$$\Phi\left(z_{\text{кр}}\right) = \frac{1 - \alpha}{2} = \frac{0{,}95}{2} = 0{,}475$$

по таблице (см. приложение, табл. 2) находим $z_{\text{кр}} = 1{,}96$. Далее вычисляем $\frac{z_{\text{кр}} \cdot \sigma}{\sqrt{n}} = \frac{1{,}96 \cdot 12}{\sqrt{36}} \approx 3{,}92$, следовательно, доверительный интервал имеет вид

$$\left(\bar{x} - \frac{z_{\text{кр}} \cdot \sigma}{\sqrt{n}};\ \bar{x} + \frac{z_{\text{кр}} \cdot \sigma}{\sqrt{n}} \right) = (41{,}08;\ 48{,}92).$$

Пример 103. Найти доверительный интервал с надежностью 0,95 для оценки неизвестного математического ожидания a нормально распределенной случайной величины X, если дисперсия этой случайной величины $\sigma^2 = 16$, выборочное среднее $\bar{x}_n{}^2 = 16$, а объём выборки $n = 25$.

Решение. Нужно определить доверительный интервал

$$\bar{x}_n - t\frac{\sigma}{\sqrt{n}} < a < \bar{x}_n + t\frac{\sigma}{\sqrt{n}}$$

Для этого вычислим значения t и σ. Из соотношения

$$\Phi(t) = \frac{\alpha}{2} = \frac{0{,}95}{2} = 0{,}475$$

По прил. 1 находим

$$t = 1,96.$$

Среднее квадратическое отклонение

$$\sigma = \sqrt{\sigma^2} = \sqrt{16} = 4.$$

Подставив все значения получим искомый доверительный интервал:

$$15 - 1,96 \cdot \frac{4}{\sqrt{25}} < a < 15 + 1,96 \cdot \frac{4}{\sqrt{25}}.$$

т.е.

$$13,432 < a < 16,568.$$

Пример 104. С конвейера поступают электрические лампы. Каким должен быть минимальный размер партии электроламп для того, чтобы с надежностью 0,99 точность оценки математического ожидания a случайной величины X, которая характеризует продолжительность горения лампы, по выборочному среднему составляла $\varepsilon = 1$ч, если известно среднее квадратическое отклонение случайной величины $X : \sigma = 3$ ч ?
Считается, что случайная величина X имеет нормальный закон распределения.

Решение. Поскольку доверительный интервал для математического ожидания a случайной величины X вычисляется по формуле

$$\overline{x}_n - t\frac{\sigma}{\sqrt{n}} < a < \overline{x}_n + t\frac{\sigma}{\sqrt{n}},$$

точность ε оценки определяется так:

$$\varepsilon = t\frac{\sigma}{\sqrt{n}}.$$

Отсюда получаем формулу для вычисления минимального обьёма выборки, который обеспечивает заданную точность оценивания:

$$n = \min\left\{n : n > t^2 \frac{\sigma^2}{\varepsilon^2}\right\}.$$

Найдём t из соотношения $\Phi(t) = \frac{0,99}{2} = 0,495$ по прил. 1: $t = 2,58$.

Итак, минимальный объём выборки

$$n = \min\left\{n : n > 2,582 \cdot \frac{32}{12} = 23.22\right\} = 24 .$$

2.2.11. Доверительный интервал для математического ожидания при неизвестной дисперсии генеральной совокупности

Пусть случайная величина X распределена нормально, причём среднее квадратическое отклонение σ этого распределения неизвестно. Требуется построить доверительный интервал для неизвестного математического ожидания μ с заданным уровнем значимости α. Как показано ранее, случайная величина $t = \dfrac{(\bar{x} - \mu)\sqrt{n}}{\hat{S}}$ распределена по закону Стьюдента, поэтому, выбрав вероятность p и зная объём выборки n, можно по таблице найти такое $t_{n,\,p}$, что $p\left(\left|\dfrac{(\bar{x} - \mu)\sqrt{n}}{\hat{S}}\right| < t_{n,\,p}\right) = 1 - \alpha$. Проведём преобразование формулы, позволяющее оценить

$$\mu: \quad p\left(|\bar{x} - \mu| < \frac{t_{n,\,p} \cdot \hat{S}}{\sqrt{n}}\right) < 1 - \alpha$$

или

$$p\left(-\frac{t_{n,\,p} \cdot \hat{S}}{\sqrt{n}} < \bar{x} - \mu < \frac{t_{n,\,p} \cdot \hat{S}}{\sqrt{n}}\right) < 1 - \alpha,$$

откуда

$$p\left(\bar{x} - \frac{t_{n,\,p} \cdot \hat{S}}{\sqrt{n}} < \mu < \bar{x} + \frac{t_{n,\,p} \cdot \hat{S}}{\sqrt{n}}\right) = 1 - \alpha.$$

Поэтому с вероятностью (надёжностью) $p = 1 - \alpha$ можно утверждать, что интервал $\left(\bar{x} - \dfrac{t_{n,\,p} \cdot \hat{S}}{\sqrt{n}} < \mu < \bar{x} + \dfrac{t_{n,\,p} \cdot \hat{S}}{\sqrt{n}}\right)$ является доверительным для оценки неизвестного математического ожидания μ.

Пример 105. Требуется построить доверительный интервал для оценки неизвестного математического ожидания μ при $n = 9, \quad p = 0{,}95, \quad \bar{x} = 6, \quad \hat{S} = 3.$

Решение. По таблице (см. приложение, табл. 3) значениям $n = 9, \quad p = 0{,}95$ соответствует $t_{9;\,0{,}95} = 2{,}31$, поэтому

$$\left(\bar{x} - \frac{t_{n,\,p} \cdot \hat{S}}{\sqrt{n}} < \mu < \bar{x} + \frac{t_{n,\,p} \cdot \hat{S}}{\sqrt{n}} \right) =$$

$$= \left(6 - \frac{3 \cdot 2{,}31}{\sqrt{9}} < \mu < 6 + \frac{3 \cdot 2{,}31}{\sqrt{9}} \right) = (3{,}69 < \mu < 8{,}31).$$

2.2.12. Доверительный интервал для дисперсии

Пусть случайная величина X распределена нормально. Требуется построить доверительный интервал для дисперсии генеральной совокупности σ^2 либо по выборочной дисперсии S_x^2, либо по $\hat{S}^2$. То есть два случая: 1) математическое ожидание генеральной совокупности известно, 2) математическое ожидание генеральной совокупности неизвестно.

Построение доверительного интервала для дисперсии основывается на том, что случайная величина $\frac{n \cdot S_x^2}{\sigma^2}$ имеет распределение χ^2 с $k = n$ степенями свободы, величина $\frac{n \cdot \hat{S}^2}{\sigma^2}$ имеет распределение χ^2 с $k = n - 1$ степенями свободы.

Подробно рассмотрим построение доверительного интервала для второго случая, так как именно он наиболее часто встречается на практике. Итак, для выбранной вероятности $p = 1 - \alpha$, учитывая, что $\frac{n \cdot \hat{S}^2}{\sigma^2}$ имеет распределение χ^2 с $k = n - 1$ степенями свободы, можно записать $p\left(\chi_1^2 < \frac{n \cdot \hat{S}^2}{\sigma^2} < \chi_2^2 \right) = 1 - \alpha$.

Далее по таблице распределения χ^2 (см. приложение, табл. 4) нужно выбрать два значения χ_1^2 и χ_2^2, чтобы площадь, заключённая под дифференциальной функцией распределения χ^2 между χ_1^2 и χ_2^2, была равна $(1-\alpha)$. Обычно χ_1^2 и χ_2^2 выбирают такими, чтобы $p\left(\chi^2 < \chi_1^2\right) = p\left(\chi^2 > \chi_2^2\right) = \frac{\alpha}{2}$. Так как таблица содержит $p\left(\chi^2 > \chi_k^2\right)$, то $p\left(\chi^2 < \chi_1^2\right) = 1 - p\left(\chi^2 > \chi_1^2\right) = 1 - \frac{\alpha}{2}$.

Определив по таблице χ_1^2 и χ_2^2, запишем неравенство $\chi_1^2 < \frac{n \cdot \hat{S}^2}{\sigma^2} < \chi_2^2$.

Запишем неравенство, обратное данному, тогда знаки неравенства изменятся на противоположные $\frac{1}{\chi_1^2} > \frac{\sigma^2}{n \cdot \hat{S}^2} > \frac{1}{\chi_2^2}$ или $\frac{1}{\chi_2^2} < \frac{\sigma^2}{n \cdot \hat{S}^2} < \frac{1}{\chi_1^2}$. Умножая обе части неравенства на $n \cdot \hat{S}^2 > 0$, окончательно получаем доверительный интервал для дисперсии генеральной совокупности $\frac{n \cdot \hat{S}^2}{\chi_2^2} < \sigma^2 < \frac{n \cdot \hat{S}^2}{\chi_1^2}$.

Пример 106. Построить доверительный интервал с вероятностью $p = 0{,}95$ для дисперсии генеральной совокупности случайной величины X, распределённой нормально, если $\hat{S}^2 = 10, \quad n = 20$.

Решение. Доверительная вероятность $p = 1 - \alpha = 0{,}95$, тогда $\alpha = 0{,}05$. По таблице (см. приложение, табл. 4) находим значения χ_1^2 и χ_2^2. Для $p_2 = \frac{\alpha}{2} = 0{,}025 \quad k = n - 1 = 19 \quad \chi_2^2 = 3{,}29 \cdot$ Для $p_1 = 1 - \frac{\alpha}{2} = 0{,}975 \quad k = n - 1 = 19 \quad \chi_1^2 = 8{,}91 \cdot$

Тогда доверительный интервал имеет вид $\frac{20 \cdot 10}{3{,}29} < \sigma^2 < \frac{20 \cdot 10}{8{,}91}$ или $6{,}079 < \sigma^2 < 22{,}45$. Для оценки среднего квадратического отклонения доверительный интервал $\sqrt{6{,}079} < \sigma < \sqrt{22{,}45}$ или $2{,}466 < \sigma < 4{,}738$.

2.2.13. Понятие статистической гипотезы

Общая постановка задачи проверки гипотез. Под статистической гипотезой понимают всякое высказывание о генеральной совокупности, проверяемое по выборке. Статистические гипотезы подразделяют на гипотезы о законах распределения и гипотезы о параметрах распределения.

Статистическая гипотеза называется непараметрической, если в ней сформулировано предположение относительно функции распределения.

Статистическая гипотеза называется параметрической, если в ней сформулировано предположение относительно значений параметров функции распределения известного вида.

Наиболее полное и безошибочное суждение относительно истинности такого вида гипотез можно было бы сделать при исследовании всей генеральной совокупности. Однако на практике сплошное исследование по ряду причин провести невозможно. Таким образом, суждения об истинности (ложности) статистических гипотез относительно вида функции распределения генеральной совокупности $F(x;\,\theta)$ или о значениях параметров распределения известного вида принимаются на основании выборки объёма n. Процесс использования выборки для проверки истинности (ложности) статистических гипотез называется статистическим доказательством истинности (ложности) выдвинутой гипотезы. Наряду с выдвинутой гипотезой рассматривают одну или несколько альтернативных (конкурирующих) гипотез. Если выдвинутая гипотеза будет отвергнута, то её место займёт конкурирующая гипотеза. С этой точки зрения статистические гипотезы подразделяются на нулевые и альтернативные.

Нулевой гипотезой называют основную (выдвинутую) гипотезу. Нулевую гипотезу обозначают символом H_0. Обычно нулевые гипотезы утверждают, что различие между сравниваемыми величинами (параметрами или функциями распределения) отсутствует, а наблюдаемое отклонение объясняется лишь случайными колебаниями выборки.

Альтернативной называется гипотеза, конкурирующая с нулевой гипотезой в том смысле, что если нулевая гипотеза отвергается, то принимается альтернативная. Альтернативную гипотезу обозначают символом H_1.

Параметрическая гипотеза называется простой, если она содержит только одно предположение относительно параметра.

Параметрическая гипотеза называется сложной, если она состоит из конечного или бесконечного числа простых гипотез.

Проверка статистических гипотез осуществляется на основе данных выборки. Для этого используют специальным образом подобранную случайную величину (выборочную статистику), являющуюся функцией наблюдённых значений, точное или приближённое распределение которой известно.

Статистическим критерием (тестом) называют случайную величину K, с помощью которой принимаются решения о принятии или отвержении выдвинутой нулевой гипотезы.

Для проверки нулевых гипотез по выборочным данным вычисляют частные значения входящих в критерий величин и, таким образом, получают частное (наблюдаемое) значение критерия.

Проверка гипотезы с помощью статистического критерия значимости есть правило отклонения нулевой гипотезы, заключающееся в разбиении области возможных значений K на две непересекающиеся подобласти, причём нулевая гипотеза отвергается, если наблюдаемое значение критерия K принадлежит критической подобласти, и считается согласующейся с опытом, если наблюдаемое значение критерия K не принадлежит критической подобласти.

2.2.14. Ошибки, допускаемые при проверке статистических гипотез

Уровень значимости статистического критерия. Ошибкой первого рода называется ошибка отклонения верной нулевой гипотезы H_0.

Уровнем значимости статистического критерия называется вероятность α совершения ошибки первого рода.

Отклонение нулевой H_0 гипотезы при уровне значимости $\alpha = 0{,}05$ означает, что, отклоняя эту гипотезу, мы или не ошибаемся (то есть гипотеза H_0 действительно ложная), или всё-таки совершаем ошибку первого рода, считая правильную гипотезу H_0 ложной.

Ошибкой второго рода называется ошибка принятия ложной гипотезы H_0.

Вероятность совершения ошибки второго рода принято обозначать β.

Мощностью M критерия K называется вероятность $(1-\beta)$ несовершения ошибки второго рода (мощность критерия K – это вероятность отклонения неверной гипотезы H_1).

2.2.15. Проверка гипотезы о равенстве математических ожиданий двух нормальных генеральных совокупностей при известной дисперсии

Проверка гипотез о равенстве двух центров распределения имеет важное практическое значение. Действительно, иногда оказывается, что средний результат одной серии экспериментов заметно отличается от среднего результата другой серии. При этом возникает вопрос: можно ли объяснить обнаруженное расхождение средних случайными ошибками экспериментов или оно вызвано какими-либо незамеченными или даже неизвестными закономерностями.

Сформулируем задачу сравнения двух центров распределения в общем виде. Рассмотрим две случайные величины X и Y, каждая из которых подчиняется нормальному закону распределения. Пусть имеются две независимые выборки объёмами n_1 и n_2 из генеральных совокупностей X и Y. Необходимо проверить гипотезу H_0, заключающуюся в том, что $M(x) = M(y)$, относительно альтернативной гипотезы H_1, состоящей в том, что $M(x) \neq M(y)$. Так как в рассматриваемом случае дисперсии генеральных совокупностей σ_x^2 и σ_y^2 известны, а о значениях математических ожиданий $M(x)$ и $M(y)$ ничего неизвестно, то для проверки гипотезы H_0 используем их оценки $\bar{x}$ и $\bar{y}$. Как известно, выборочные средние $\bar{x}$ и $\bar{y}$ имеют нормальный закон распределения с параметрами $\left(M(x); \dfrac{\sigma_x^2}{n_1}\right)$ и $\left(M(y); \dfrac{\sigma_y^2}{n_2}\right)$.

Выборки независимы, поэтому $\bar{x}$ и $\bar{y}$ также независимы и случайная величина, равная разности $(\bar{x}-\bar{y})$, имеет нормальное распределение, причём $D(\bar{x}-\bar{y}) = D(\bar{x}) + D(\bar{y}) = \dfrac{\sigma_x^2}{n_1} + \dfrac{\sigma_y^2}{n_2}$. Если гипотеза H_0 справедлива, то $M(\bar{x}-\bar{y}) = M(\bar{x}) - M(\bar{y}) = 0$, следовательно, нормированная разность $z = \dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\sigma_x^2}{n_1} + \dfrac{\sigma_y^2}{n_2}}}$ подчиняется нормальному закону с математическим ожиданием, равным нулю, и дисперсией, равной единице. Выбирая вероятность $p = 1-\alpha$ из соотношения $\Phi(z_{\text{кр}}) = \dfrac{1-\alpha}{2}$, находим статистику $z_{\text{кр}}$. Если $|z_{\text{наб}}| \le z_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу или при заданном уровне значимости можно считать, что $M(x) = M(y)$. Если $|z_{\text{наб}}| > z_{\text{кр}}$, то нулевую гипотезу отвергаем в пользу конкурирующей или при заданном уровне значимости можно считать, что математические ожидания генеральных совокупностей различны.

2.2.16. Сравнение выборочных средних при неизвестной дисперсии генеральной совокупности

Рассмотрим две случайные величины X и Y, каждая из которых подчиняется нормальному закону распределения с математическими ожиданиями $M(x)$ и $M(y)$. Условимся считать, что $\sigma_x^2 = \sigma_y^2 = \sigma^2$, числовое значение σ^2 неизвестно. Пусть имеются две независимые выборки объёмами n_1 и n_2 из генеральных совокупностей X и Y. Необходимо проверить гипотезу H_0, заключающуюся в том, что $M(x) = M(y)$ относительно альтернативной гипотезы H_1, состоящей в том, что $M(x) \neq M(y)$. Для оценки $M(x)$ и $M(y)$ используем их оценки $\bar{x}$ и $\bar{y}$, а для оценки σ^2 используем выборочные оценки: $\hat{S}_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2$ и $\hat{S}_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(y_i - \bar{y})^2$. Так как генеральные совокупности X и Y имеют одинаковые дисперсии, то для оценки σ^2 целесообразно использовать результаты обеих выборок. В математической статистике доказано, что лучшей оценкой для σ^2 в данном случае является $\hat{S}^2 = \frac{(n_1 - 1)\cdot \hat{S}_1^2 + (n_2 - 1)\cdot \hat{S}_2^2}{n_1 - 1 + n_2 - 1}$. В качестве выборочной оценки $D(\bar{x} - \bar{y})$ обычно принимают оценку $\hat{S}^2_{(\bar{x}-\bar{y})} = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\cdot \hat{S}^2$. Известно, что если случайная величина $(\bar{x} - \bar{y})$ подчиняется нормальному закону, то статистика

$$t = \frac{(\bar{x} - \bar{y}) - M(\bar{x} - \bar{y})}{\hat{S}_{(\bar{x}-\bar{y})}} = \frac{(\bar{x} - \bar{y}) - M(\bar{x} - \bar{y})}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\cdot \frac{(n_1 - 1)\cdot \hat{S}_1^2 + (n_2 - 1)\cdot \hat{S}_2^2}{n_1 + n_2 - 2}}}$$

имеет t-распределения Стьюдента с $k = n_1 + n_2 - 2$ степенями свободы. Если гипотеза H_0 справедлива, то статистику t можно записать в виде: $t = \frac{(\bar{x} - \bar{y})}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\cdot \frac{(n_1 - 1)\cdot \hat{S}_1^2 + (n_2 - 1)\cdot \hat{S}_2^2}{n_1 + n_2 - 2}}}$. Выбрав вероятность $p = 1 - \alpha$, по таблице t-распределения можно определить критическое значение

$t_{n_1+n_2-2;\,\alpha}$. Если вычисленное значение по модулю меньше критического, то с надёжностью $p = 1-\alpha$ можно считать расхождение средних значимым (неслучайным).

Пример 107. В результате двух серий измерений с количеством измерений $n_1 = 25$ и $n_2 = 50$ получены следующие выборочные средние: $\overline{x} = 9{,}79$ и $\overline{y} = 9{,}6$, а также исправленные дисперсии: $\hat{S}_x^2 = 0{,}28$ и $\hat{S}_y^2 = 0{,}33$. Можно ли с надёжностью $p = 0{,}99$ объяснить расхождение между выборочными средними случайными причинами?

Решение. Вычислим исправленную дисперсию

$$\hat{S} = \sqrt{\frac{(n_1-1)\cdot \hat{S}_x^2 + (n_2-1)\cdot \hat{S}_y^2}{n_1-1+n_2-1}} = \sqrt{\frac{24\cdot 0{,}28 + 49\cdot 0{,}33}{73}} \approx 0{,}56,$$

вычислим наблюдаемое значение статистики

$$t = \frac{\left(\overline{x}-\overline{y}\right)}{\hat{S}\cdot\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{9{,}79-9{,}6}{0{,}56\cdot\sqrt{\frac{1}{25}+\frac{1}{49}}} \approx 1{,}38.$$

Вероятности $p = 0{,}99$ и числу степеней свободы $k = 73$ в таблице t-распределения (см. приложение, табл. 3) соответствует $t_{73;\,0{,}01} = 2{,}649$. Так как $1{,}38 < 2{,}649$, то с надёжностью $0{,}99$ нельзя считать расхождение средних значимым, или при уровне значимости $0{,}99$ можно считать, что математические ожидания $M(x) = M(y)$.

2.2.17. Сравнение выборочных дисперсий

Гипотезы о дисперсиях имеют особенно большое значение в технике, так как измеряемая дисперсией величина рассеяния характеризует такие важные показатели, как точность машин, приборов, технологических процессов.

Сформулируем гипотезу о равенстве дисперсий двух нормально распределённых генеральных совокупностей. Рассмотрим две случайные величины X и Y, каждая из которых подчиняется нормальному закону распределения с дисперсиями σ_x^2 и σ_y^2. Пусть из генеральных совокупностей X и Y извлечены две независимые выборки объёмами n_1 и n_2. Проверим гипотезу H_0 о том, что $\sigma_x^2 = \sigma_y^2$ относительно альтернативной гипотезы H_1, заключающейся в том, что $\sigma_x^2 > \sigma_y^2$. Для оценки σ_x^2 используем исправленную выборочную дисперсию $\hat{S}_x^2$, а для оценки σ_y^2 – исправленную выборочную дисперсию $\hat{S}_y^2$, следовательно, задача проверки гипотезы H_0 сводится к сравнению дисперсий $\hat{S}_x^2$ и $\hat{S}_y^2$. Как показано ранее, случайные величины $\dfrac{\hat{S}_x^2 \cdot n_2}{\sigma^2}$ и $\dfrac{\hat{S}_y^2 \cdot n_2}{\sigma^2}$ распределены по закону χ^2 с $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$ степенями свободы. Случайную величину F, определяемую соотношением $F = \dfrac{n_2}{n_1} \cdot \dfrac{\dfrac{\hat{S}_1^2 \cdot n_1}{\sigma^2}}{\dfrac{\hat{S}_2^2 \cdot n_2}{\sigma^2}} = \dfrac{\hat{S}_1^2}{\hat{S}_2^2}$, называют случайной величиной с распределением Фишера – Снедекора. Заметим, что всегда можно так ввести обозначения, что $\hat{S}_1^2 \geq \hat{S}_2^2$, поэтому случайная величина F принимает значения, не меньшие единицы. Дифференциальный закон распределения случайной величины F не содержит неизвестных параметров $(\mu; \sigma)$ и их оценок, а зависит лишь от числа наблюдений в выборках n_1 и n_2. Этот факт позволяет составить таблицы распределения случайной величины F, в которых различным значениям уровня значимости и различным сочетаниям величин k_1 и k_2 ставят в соответствие такие значения $F(\alpha; k_1; k_2)$, для которых справедливо равенство $p(F > F(\alpha; k_1; k_2)) = \alpha$.

Сформулируем правило, вычислив исправленные выборочные дисперсии $\hat{S}_x^2$ и $\hat{S}_y^2$, найдём их отношение, причём в числителе запишем большую из них. Затем, выбрав необходимый уровень значимости α, по таблице F-распределения находим число $F(\alpha; k_1; k_2)$, которое сравниваем с вычисленным F_p. Если окажется, что $F > F(\alpha; k_1; k_2)$, то проверяемая гипотеза отвергается (различие между дисперсиями значимо), если $F \leq F(\alpha; k_1; k_2)$, то выборочные наблюдения не противоречат проверяемой гипотезе.

Пример 108. На двух станках обрабатываются детали. Отобраны две пробы: на первом станке $n_1 = 10$, на втором станке $n_2 = 15$. По данным этих выборок рассчитаны исправленные выборочные дисперсии $\hat{S}_1^2 = 9{,}6$ и $\hat{S}_2^2 = 4{,}8$. Проверить при уровне значимости $\alpha = 0{,}05$ гипотезу о том, что станки обладают одинаковой точностью, или гипотезу H_0 : дисперсии равны.

Решение. Вычислим значение $F_p = \dfrac{\hat{S}_1^2}{\hat{S}_2^2} = \dfrac{9{,}6}{4{,}8} = 2$. Затем по уровню значимости $\alpha = 0{,}05$ и степеням свободы $k_1 = n_1 - 1 = 10 - 1 = 9$ и $k_2 = n_2 - 1 = 15 - 1 = 14$ по таблице (см. приложение, табл. 6) находим число $F(9; 14; 0{,}05) = 2{,}65$. Итак, имеем $2 < 2{,}65$, следовательно, предположение о равенстве дисперсий не противоречит наблюдениям, иными словами, нет оснований считать, что станки обладают разной точностью.

Пример 109. Из нормальной генеральной совокупности получена выборка объёма $n = 25$, для которой найдена выборочная “исправленная” дисперсия $s^2 = 12{,}3$. Проверить при уровне значимости $\alpha = 0{,}01$основную гипотезу $H_0 : \sigma^2 = \sigma_0^2 = 12$, если альтернативная гипотеза $H_1 : \sigma_0^2 > 12$.

Решение. Найдём наблюдаемое значение критерия:

$$X^2_{набл} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(25-1)\cdot 12{,}3}{12} = 24{,}6 .$$

Поскольку альтернативная гипотеза имеет вид $H_1:\sigma_0^2>12$, критическая область правосторонняя (правило 1.). По прил. 7 при уровне значимости $\alpha=0{,}01$ и числе степеней свободы $k=n-1=25-1=24$ находим критическую точку:

$$X^2_{кр}(0{,}01;24)=42{,}98\ .$$

Поскольку $X^2_{набл}<X^2_{кр}$, нет оснований отклонять основную гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2\le 12$. Другими словами, разница между “исправленной” выборочной дисперсией $s^2=12{,}3$ и гипотетической генеральной дисперсией $\sigma_0^2\le 12$ незначащая.

Пример 110. Из нормальной генеральной совокупности получена выборка объёма $n=20$, для которой найдена выборочная “исправленная” дисперсия $s^2=2{,}7$. Проверить при уровне значимости $\alpha=0{,}05$основную гипотезу $H_0:\sigma^2=\sigma_0^2=3$если альтернативная гипотеза:

а) $H_1:\sigma_0^2>3$;

б) $H_1:\sigma_0^2\ne 3$;

в) $H_1:\sigma_0^2<3$.

Решение. Найдём наблюдаемое значение критерия:

$$X^2_{набл}=\frac{(n-1)s^2}{\sigma_0^2}=\frac{(20-1)\cdot 2{,}7}{3}=17{,}1$$

а) Воспользуемся правилом 1. По прил. 7 при уровне значимости $\alpha=0{,}05$ и числе степеней свободы $k=n-1=20-1=19$ находим критическую точку

$$X^2_{кр}(0{,}05;19)=30{,}14$$

Поскольку $X^2_{набл}<X^2_{кр}$, нет оснований отклонят основную гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2\le 3$. Другими словами, разница между “исправленной” выборочной дисперсией $s^2=2{,}7$ и гипотетической генеральной дисперсией $\sigma_0^2\le 3$ незначащая.

б) Воспользуемся правилом 2. По прил. 7 при уровне значимости $\alpha = 0,05$ и числе степеней свободы $k = n - 1 = 20 - 1 = 19$. Находим критические точки:

$$X^2_{лев.кр}\left(1 - \alpha / 2; k\right) = X^2_{лев.кр}\left(0,975; 19\right) = 8,91$$

$$X^2_{прав.кр}\left(\alpha / 2; k\right) = X^2_{прав.кр}\left(0,025; 19\right) = 32,85 .$$

Поскольку $X^2_{лев.кр} < X^2_{набл} < X^2_{прав.кр}$, нет оснований отклонять основную гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2 = 3$. Другими словами, разница между "исправленной" выборочной дисперсией $s^2 = 2,7$ и гипотетической генеральной дисперсией $\sigma_0^2 = 3$ незначащая.

в) Воспользуемся правилом 3. По прил. 7 при уровне значимости $\alpha = 0,05$ и числе степеней свободы $k = n - 1 = 20 - 1 = 19$. Находим критические точки:

$$X^2_{кр}\left(1 - \alpha; k\right) = X^2_{кр}\left(0,95; 19\right) = 10,12 .$$

Поскольку $X^2_{набл} > X^2_{кр}\left(1 - \alpha; k\right)$, нет оснований отклонять основную гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2 \geq 3$. Другими словами , разница между "исправленной" выборочной дисперсией $s^2 = 2,7$ и гипотетической генеральной дисперсией $\sigma_0^2 \geq 3$ незначащая.

Пример 111. Из нормальной генеральной совокупности получена выборка объёма $n = 100$:

x_i	5	6	7	8	9	10	11
n_i	9	15	19	20	17	13	7

Необходимо проверить при уровне значимости $\alpha = 0,01$ основную гипотезу $H_0 : \sigma^2 = \sigma_0^2 = 4$, если альтернативная гипотеза $H_1 : \sigma_0^2 > 4$.

Решение. Сначала по данным выборки найдём "исправленную" выборочную дисперсию s^2:

$$\bar{x} = \frac{5 \cdot 9 + 6 \cdot 15 + 7 \cdot 19 + 8 \cdot 20 + 9 \cdot 17 + 10 \cdot 13 + 11 \cdot 7}{100} = 7,88 ;$$

$$s^2 = \frac{5^2 \cdot 9 + 6^2 \cdot 15 + 7^2 \cdot 19 + 8^2 \cdot 20 + 9^2 \cdot 17 + 10^2 \cdot 13 + 11^2 \cdot 7}{100} - 7,88^2 = 2,9056 .$$

Вычислим наблюдаемое значение критерия:

$$X^2_{набл} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(100-1)\cdot 2,9056}{4} = 71,9136 \ .$$

Поскольку альтернативная гипотеза имеет вид $H_1 : \sigma_0^2 > 12$, критическая область правосторонняя (правило 1).

Найдем число степеней свободы к: $k = n - 1 = 100 - 1 = 99$.

Поскольку число степеней свободы $k > 30$, критическую точку $X^2_{кр}(\alpha; k)$ найдём из равенства Уилсона-Гильферти:

$$X^2_{кр}(\alpha; k) = k\left(1 - \frac{2}{9k} + z_\alpha\sqrt{\frac{2}{9k}}\right)^3 ,$$

где z_α определим из равенства

$$\Phi(z_\alpha) = \frac{1}{2} - \alpha = \frac{1}{2} - 0,01 = 0,49 ,$$

Используя таблицу функции Лапласа (прил. 1):

$$z_\alpha \approx 2,33.$$

Итак ,

$$X^2_{кр}(0,01; 99) = 99\left(1 - \frac{2}{9\cdot 99} + 2,33\sqrt{\frac{2}{9\cdot 99}}\right)^3 = 113,0187 \ .$$

Поскольку $X^2_{набл} < X^2_{кр}$, нет оснований отклонять основную гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2 \le 4$. Другими словами , разница между “исправленной” выборочной дисперсией $s^2 \approx 2,9$ и гипотетической генеральной дисперсией $\sigma_0^2 \le 4$ незначащая.

2.2.18. Проверка гипотез о законе распределения. Критерий согласия χ^2 (Пирсона)

До сих пор мы рассматривали гипотезы, относящиеся к отдельным параметрам распределения случайной величины, причём закон распределения предполагался известным. Однако во многих практических задачах закон распределения случайной величины неизвестен, то есть является гипотезой, которая требует статистической проверки.

Обозначим через X исследуемую случайную величину. Пусть требуется проверить гипотезу H_0 о том, что случайная величина подчиняется закону распределения $F(x)$. Для проверки гипотезы произведём выборку, состоящую из n независимых наблюдений над случайной величиной X. По выборке можно построить эмпирическое распределение $F^*(x)$ исследуемой случайной величины. Сравнение эмпирического $F^*(x)$ и теоретического распределений производится с помощью специально подобранной случайной величины – критерия согласия. Существует несколько критериев согласия: χ^2 Пирсона, Колмогорова, Смирнова и др. Критерий согласия Пирсона – наиболее часто употребляемый критерий для проверки гипотезы о законе распределения.

Рассмотрим этот критерий. Разобьём всю область изменения X на l интервалов $\Delta_1, \Delta_2, ..., \Delta_l$ и подсчитаем количество элементов m_i, попавших в каждый из интервалов Δ_i. Предполагая известным теоретический закон распределения $F(x)$, всегда можно определить p_i – вероятность попадания случайной величины X в интервал Δ_i, тогда теоретические частоты можно рассчитать по формуле $n \cdot p_i$. Если эмпирические частоты сильно отличаются от теоретических, то проверяемую гипотезу H_0 следует отвергнуть, в противном случае – принять.

Сформулируем критерий, который бы характеризовал степень расхождения между эмпирическими и теоретическими частотами. В литературе по математической статистике доказывается, что статистика $\chi^2 = \sum\limits_{i=1}^{n} \frac{(m_i - np_i)}{np_i}$ имеет распределение χ^2 с $k = l - r - 1$ степенями свободы. Здесь r – число параметров распределения $F(x)$. Правило применения критерия χ^2 сводится к следующему. Рассчитав теоретические частоты np_i и вычислив значение χ_p^2, затем выбрав уровень значимости критерия α, по таблице находим $\chi^2(k;\alpha)$. Если $\chi_p^2 > \chi^2(k;\alpha)$, то гипотезу H_0 отвергают; если $\chi_p^2 \le \chi^2(k;\alpha)$, то гипотезу принимают или, другими словами, нет оснований отвергнуть гипотезу о том, что генеральная совокупность подчинена закону распределения $F(x)$. В заключение отметим, что необходимым условием применения критерия Пирсона является наличие в каждом из интервалов по меньшей мере 5–10 наблюдений. Если количество наблюдений мало, то нужно объединить интервалы, содержащие частоты менее 5.

Пример 112. На телефонной станции производились наблюдения над числом X неправильных соединений в минуту. Наблюдения в течение часа дали следующие результаты: 3; 1; 3; 1; 4; 2; 2; 4; 0; 3; 0; 2; 2; 0; 2; 1; 4; 3; 3; 1; 4; 2; 2; 1; 1; 2; 1; 0; 3; 4; 1; 3; 2; 7; 2; 0; 0; 1; 3; 3; 1; 2; 4; 2; 0; 2; 3; 1; 2; 5; 1; 1; 0; 1; 1; 2; 2; 1; 1; 5. Определить выборочную среднюю и дисперсию неправильных соединений в минуту и проверить выполнение основного условия для распределения Пуассона $M(x) = \sigma^2$. Найти теоретическое распределение Пуассона и проверить степень согласия теоретического и эмпирического распределений по критерию χ^2 Пирсона при уровне значимости $\alpha = 0{,}05$.

Решение. Упорядочим результаты наблюдений, записав их в таблицу:

x_i	0	1	2	3	4	5	6	7
m_i	8	17	16	10	6	2	0	1

Найдём $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot m_i$.

$$\bar{x} = \frac{1}{60}(0\cdot 8 + 1\cdot 17 + 2\cdot 16 + 3\cdot 10 + 4\cdot 6 + 5\cdot 2 + 6\cdot 0 + 7\cdot 1) = 2.$$

Вычислим выборочную дисперсию по формуле

$$S^2 = \overline{x^2} - \overline{x}^2,$$

где

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \cdot m_i =$$

$$= \frac{1}{60}\left(0^2 \cdot 8 + 1^2 \cdot 17 + 2^2 \cdot 16 + 3^2 \cdot 10 + 4^2 \cdot 6 + 5^2 \cdot 2 + 6^2 \cdot 0 + 7^2 \cdot 1\right) = 6{,}1.$$

Тогда $S^2 = \overline{x^2} - \overline{x}^2 = 6{,}1 - 2^2 = 2{,}1$.

Необходимое условие для распределения Пуассона выполняется. Запишем теоретический закон распределения Пуассона, используя вместо математического ожидания его оценку $m = \overline{x}$: $p(m) = \frac{2^m}{m!} \cdot e^{-2}$. Найдём теоретические частоты:

$$m'_0 = n \cdot p(0) = 60 \cdot \frac{2^0}{0!} \cdot e^{-2} \approx 8{,}1;$$

$$m'_1 = n \cdot p(1) = 60 \cdot \frac{2^1}{1!} \cdot e^{-2} \approx 16{,}24;$$

$$m'_2 = n \cdot p(2) = 60 \cdot \frac{2^2}{2!} \cdot e^{-2} \approx 16{,}24;$$

$$m'_3 = n \cdot p(3) = 60 \cdot \frac{2^3}{3!} \cdot e^{-2} \approx 10{,}62;$$

$$m'_4 = n \cdot p(4) = 60 \cdot \frac{2^4}{4!} \cdot e^{-2} \approx 5{,}41;$$

$$m'_5 = n \cdot p(5) = 60 \cdot \frac{2^5}{5!} \cdot e^{-2} \approx 2{,}17;$$

$$m'_6 = n \cdot p(6) = 60 \cdot \frac{2^6}{6!} \cdot e^{-2} \approx 0{,}72;$$

$$m'_7 = n \cdot p(7) = 60 \cdot \frac{2^7}{7!} \cdot e^{-2} \approx 0{,}2.$$

Так как последних три интервала содержат частоты менее пяти, то объединим их с предыдущим. Получим

m_i	8	17	16	10	9
m_i'	8,10	16,24	16,24	10,62	8,50

Вычислим значение

$$\chi^2 = \sum_{i=1}^{5} \frac{(m_i - m_i')^2}{m_i'} =$$

$$= \frac{(8-8{,}1)^2}{8{,}1} + \frac{(17-16{,}24)^2}{16{,}24} + \frac{(16-16{,}24)^2}{16{,}24} +$$

$$+ \frac{(10-10{,}62)^2}{10{,}62} + \frac{(9-8{,}5)^2}{8{,}5} \approx 0{,}11.$$

Примем уровень значимости $\alpha = 0{,}05$. Количество интервалов после объединения $l = 5$. По выборке вычислен один параметр, которым определяется закон Пирсона – математическое ожидание, следовательно, $r = 1$. Поэтому число степеней свободы $k = l - r - 1 = 5 - 1 - 1 = 3$. По таблице (см. приложение, табл. 4) находим $\chi^2(3;\, 0{,}05) = 7{,}8$. Имеем $7{,}8 > 0{,}11$, следовательно, нет оснований отвергнуть нулевую гипотезу или, другими словами, при уровне значимости $\alpha = 0{,}05$ можно считать, что генеральная совокупность распределена по закону Пуассона.

Пример 113. Проверить гипотезу о нормальном распределении для следующего интервального ряда:

x_i	10–14	14–18	18–22	22–26	26–30	30–34
m_i	5	8	14	12	8	3

Решение. Найдём выборочное среднее по формуле

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \cdot m_i =$$

$$= \frac{1}{50}(12 \cdot 5 + 16 \cdot 8 + 20 \cdot 14 + 24 \cdot 12 + 28 \cdot 8 + 5 \cdot 2 + 32 \cdot 3) = 21{,}52$$

и исправленную дисперсию по формуле

$$\hat{S}^2 = \frac{n}{n-1}\left(\overline{x^2} - \overline{x}^2\right),$$

где

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \cdot m_i =$$

$$= \frac{1}{50}\left(12^2 \cdot 5 + 16^2 \cdot 8 + 20^2 \cdot 14 + 24^2 \cdot 12 + 28^2 \cdot 8 + 32^2 \cdot 3\right) = 492{,}48;$$

$$\hat{S}^2 = \frac{n}{n-1}\left(\overline{x^2} - \overline{x}^2\right) = \frac{50}{49}\left(492{,}48 - 21{,}52^2\right) \approx 29{,}969;$$

$$\hat{S} \approx 5{,}474.$$

Найдём теоретические частоты по формуле

$$m_i' = n \cdot \left(\Phi\left(\frac{x_{i+1} - \overline{x}}{\hat{S}}\right) - \Phi\left(\frac{x_i - \overline{x}}{\hat{S}}\right)\right),$$

тогда

$$m_1' = 50 \cdot \left(\Phi\left(\frac{14 - 21{,}52}{5{,}474}\right) - \Phi\left(\frac{10 - 21{,}52}{5{,}474}\right)\right) =$$

$$= 50 \cdot (-0{,}4152 + 0{,}4823) \approx 3{,}3;$$

$$m_2' = 50 \cdot \left(\Phi\left(\frac{18 - 21{,}52}{5{,}474}\right) - \Phi\left(\frac{14 - 21{,}52}{5{,}474}\right)\right) =$$

$$= 50 \cdot (-0{,}2399 + 0{,}4152) \approx 8{,}8;$$

$$m_3' = 50 \cdot \left(\Phi\left(\frac{22 - 21{,}52}{5{,}474}\right) - \Phi\left(\frac{18 - 21{,}52}{5{,}474}\right)\right) =$$

$$= 50 \cdot (0{,}0349 + 0{,}2399) \approx 13{,}7;$$

$$m'_4 = 50 \cdot \left(\Phi\left(\frac{26 - 21{,}52}{5{,}474} \right) - \Phi\left(\frac{22 - 21{,}52}{5{,}474} \right) \right) =$$

$$= 50 \cdot (0{,}2399 - 0{,}0349) \approx 12{,}9;$$

$$m'_5 = 50 \cdot \left(\Phi\left(\frac{30 - 21{,}52}{5{,}474} \right) - \Phi\left(\frac{26 - 21{,}52}{5{,}474} \right) \right) =$$

$$= 50 \cdot (0{,}4393 - 0{,}2399) \approx 7{,}3;$$

$$m'_6 = 50 \cdot \left(\Phi\left(\frac{34 - 21{,}52}{5{,}474} \right) - \Phi\left(\frac{30 - 21{,}52}{5{,}474} \right) \right) =$$

$$= 50 \cdot (0{,}4887 - 0{,}24393) \approx 2{,}5.$$

Так как последний и первый интервалы содержат частоты менее пяти, то объединим их с соседними.

m_i	13	14	12	11
m'_i	12,1	13,7	12,9	9,8

Найдём:Число степеней свободы равно $k = l - r - 1 = 4 - 2 - 1 = 1$, следовательно, при уровне значимости $\alpha = 0{,}05$ имеем по таблице $\chi^2_{\text{kp}}(1;\ 0{,}05) = 3{,}8$. Так как наблюдаемое значение меньше критического, то гипотезу о нормальном распределении генеральной совокупности при заданном уровне значимости нет оснований отвергнуть.

Пример 114. Используя критерий Пирсона, при уровни значимости 0,05 проверить, выполняется ли гипотеза о нормальном распределении генеральной совокупности *X,* если из этой совокупности получена такая выборка объёма $n = 100$:

x_i	-5	-3	-1	1	3	5	7	9
n_i	5	9	13	18	21	18	10	6

Решение. Сначала вычислим выборочное среднее $\overline{x}_n$ и среднее квадратическое отклонение выборки σ_n :

$$\overline{x}_n = \frac{-5\cdot 5 - 3\cdot 9 - 1\cdot 13 + 1\cdot 18 + 3\cdot 21 + 5\cdot 18 + 7\cdot 10 + 9\cdot 6}{100} = 2,3;$$

$$\sigma_n = \left(\left[5^2\cdot 5 + 3^2\cdot 9 + 1\cdot 13 + 1\cdot 18 + 3^2\cdot 21 + 5^2\cdot 18 + 7^2\cdot 10 + 9^2\cdot 6\right]/100 - 2,32\right)^{\frac{1}{2}} \approx 3,64.$$

Определим теоретические частоты, учитывая, что $n=100,\ h=2,\ \sigma_n=3,64,$ по формуле

$$n_i' = \frac{nh}{\sigma_n}\cdot\phi\left(\frac{x_i-\overline{x}_n}{\sigma_n}\right).$$

Для этого составим расчётную табл.14 (значение дифференциальной функции Лапласа $\phi(u)$ приведено в пил. 2).

Сравним эмпирические и теоретические частоты. Для этого вычислим наблюдаемое значение критерия

$$\chi^2_{набл} = \sum_i \frac{\left(n_i - n_i'\right)^2}{n_i'}$$

с помощью расчётной табл. 15.

Таблица 14

i	x_i	$u_i = \frac{x_i-\overline{x}_n}{\sigma_n}$	$\varphi(u_i)$	$n_i' = \frac{nh}{\sigma_n}\cdot\varphi\left(\frac{x_i-\overline{x}_n}{\sigma_n}\right)$
1	-5	-2,01	0,0529	2,91
2	-3	-1,46	0,1374	7,56
3	-1	-0,91	0,2637	14,50
4	1	-0,36	0,3739	20,56
5	3	0,19	0,3918	21,54
6	5	0,74	0,3034	16,68
7	7	1,29	0,1736	9,55
8	9	1,84	0,0734	4,04

Таблица 15

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
1	5	2,91	2,09	4,37	1,50
2	9	7,56	1,44	2,07	0,27
3	13	14,50	-1,50	2,25	0,16
4	18	20,56	-2,56	6,55	0,32
5	21	21,54	0,29	0,29	0,01
6	18	16,68	1,32	1,74	0,10
7	10	9,55	0,45	0,20	0,02
8	6	4,04	1,96	3,84	0,95
$\sum$	100				$\chi^2_{набл} = 3{,}33$

Из табл. 15 находим

$$\chi^2_{набл} = 3{,}33$$

По таблице критических точек распределения $\chi 2$ (прил. 7) при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $k = s - 3 = 8 - 3 = 5$ находим критическую точку правосторонней критической области: $\chi^2_{кр}(0{,}05:5) = 11{,}07$.

Поскольку $\chi^2_{набл} < \chi^2_{кр}$, гипотезу о нормальном распределении не отклоняем. Другими словами, эмпирические и теоретические частоты различаются незначаще.

Пример 115. Используя критерий Пирсона, при уровни значимости 0,05 проверить, выполняется ли гипотеза о нормальном распределении генеральной совокупности X, если из этой совокупности получена такая выборка объёма $n = 100$:

$(x_i; x_{i+1}]$	(3;7]	(7;11]	(11;15]	(15;19]	(19;23]	(23;27]	(27;31]
n_i	6	16	19	17	15	14	13

Решение. Сначала вычислим выборочное среднее $\overline{x}_n$ и среднее квадратическое отклонение выборка σ_n. для этого преобразуем интервальное статическое распределение выборки в точечное:

x_i	5	9	13	17	21	25	29
n_i	6	16	19	17	15	14	13

Итак,

$$\overline{x}_n = \frac{5\cdot 6 - 9\cdot 16 - 13\cdot 19 + 17\cdot 17 + 21\cdot 15 + 25\cdot 14 + 29\cdot 13}{100} = 17,52;$$

$$\sigma_n = \left(\left[5^2\cdot 6 + 9^2\cdot 16 + 13^2\cdot 19 + 17^2\cdot 17 + 21^2\cdot 15 + 25^2\cdot 14 + 29^2\cdot 13\right]/100 - 17,52^2\right)^{\frac{1}{2}} \approx 7,19.$$

Найдём нормированные интервалы $(z_i; z_{i+1}]$ учитывая, что выборочное среднее $\overline{x}_n = 17,52$ и среднее квадратическое отклонение выборки $\sigma_n = 7,19$. Для этого составим расчётную табл. 16 (левый конец первого интервала положим равным $-\infty$, а правый конец последнего интервала - равным ∞).

Таблица 16

i	x_i	x_{i+1}	$x_i - \overline{x}_n$	$x_{i+1} - \overline{x}_n$	$z_i = \dfrac{x_i - \overline{x}_n}{\sigma_n}$	$z_{i+1} = \dfrac{x_{i+1} - \overline{x}_n}{\sigma_n}$
1	3	7	-14,5	-10,5	$-\infty$	-1,46
2	7	11	-10,5	-6,5	-1,46	-0,90
3	11	15	-6,5	-2,5	-0,90	-0,35
4	15	19	-2,5	1,5	-0,35	0,21
5	19	23	1,5	5,5	0,21	0,76
6	23	27	5,5	9,5	0,76	1,32
7	27	31	9,5	13,5	1,32	∞

Определим теоретические вероятности P_i и теоретические частоты:

$$n_i' = n\cdot P_i.$$

Для этого составим расчётную табл. 17.

Таблица 17

	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i = \Phi(z_i) - \Phi(z_{i+1})$	$n_i' = n\cdot P_i$
1	$-\infty$	-1,48	-0,5000	-0,4306	0,0694	6,94
2	-1,48	-0,92	-0,4306	-0,3212	0,1094	10,94
3	-0,92	-0,35	-0,3212	-0,1368	0,1844	18,44
4	-0,35	0,21	-0,1368	0,0832	0,2200	22,00

5	0,21	0,78	0,0832	0,2823	0,1991	19,91
6	0,78	1,34	0,2823	0,4099	0,1276	12,76
7	1,34	∞	0,4099	0,5000	0,0901	9,01
Σ					1	100

Сравним эмпирические и теоретические частоты, испльзуя критерий Пирсона. Для этого вычислим сначала наблюдаемое значение критерия Пирсона с помощью расчётной табл.18. Столбцы 7 и 8 этой таблицы введены для контроля вычислений по формуле

$$\chi^2_{набл} = \sum_i \frac{n_i^2}{n_i^{'}} - n.$$

Таблица 18

1	2	3	4	5	6	7	8
i	n_i	$n_i^{'}$	$n_i - n_i^{'}$	$(n_i - n_i^{'})^2$	$\frac{(n_i - n_i^{'})^2}{n_i^{'}}$	n_i^2	$\frac{n_i^2}{n_i^{'}}$
1	6	6,94	-0,94	0,8836	0,1273	36	5,1873
2	16	10,94	5,06	25,6036	2,3404	256	23,4004
3	19	18,44	0,56	0,3136	0,0170	361	19,5770
4	17	22,00	-5,00	25,0000	1,1364	289	13,1364
5	15	19,91	-4,91	24,1081	1,2109	225	11,3009
6	14	12,76	1,24	1,5376	0,1205	196	15,3605
7	13	9,01	3,99	159201	1,7669	169	18,7569
Σ	100	100			$\chi^2_{набл} = 6{,}7194$		106,7194

Контроль: Поскольку

$$\sum_i \frac{n_i^2}{n_i^{'}} - n = 106,7194 - 100 = 6,7194 = \chi^2_{набл},$$

вычисление проведено правильно.

По таблице критических точек распределения χ^2 (прил. 7) при уровне значимости $\alpha = 0,05$ и числе степеней свободы $k = s - 3 = 7 - 3 = 4$ найдём критическую точку правосторонней критической области:

$$\chi^2_{кр}(0,05; 4) = 9,49.$$

Поскольку $\chi^2_{набл} < \chi^2_{кр}$, гипотезу о нормальном распределении не отклоняем. Другими словами, эмпирические и теоретические частоты различаются незначаще.

2.2.19. Выборочный коэффициент корреляции и его свойства

В реальном мире многие явления природы происходят в обстановке действия многочисленных факторов, влияние каждого из них ничтожно, а число их велико. В этом случае связь теряет свою однозначность, и изучаемая физическая система переходит не в определённое состояние, а в одно из возможных для неё состояний. Здесь речь может идти лишь о статистической связи. Знание статистической зависимости между случайными переменными имеет большое практическое значение: с её помощью можно прогнозировать значение случайной переменной в предположении, что независимая переменная примет определённое значение.

Статистические связи между переменными можно изучать методом корреляционного и регрессионного анализа. Основная задача корреляционного анализа – выявление связи между случайными переменными путём точечной и интервальной оценки парных коэффициентов корреляции, вычисления и проверки значимости коэффициентов корреляции. Корреляционный анализ позволяет оценить функцию регрессии одной случайной величины на другую.

Предпосылки корреляционного анализа следующие:

1) переменные величины должны быть случайными;
2) случайные величины должны иметь нормальное распределение.

Выборочный коэффициент корреляции находится по формуле

$$r_b = \frac{\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{n \cdot S(x) \cdot S(y)}.$$

Выборочный коэффициент корреляции оценивает тесноту линейной связи.

Свойства выборочного коэффициента корреляции:

1. Коэффициент корреляции принимает значения на интервале $[-1; 1]$.

2. Выборочный коэффициент корреляции не зависит от выбора начала точки отсчёта и единицы измерения, то есть для любых a_1, b_1, a_2, b_2 выполнено равенство:

$$r_b(a_1 \cdot x + b_1;\ a_2 \cdot y + b_2) = r_b(x;\ y).$$

3. Выборочный коэффициент можно вычислять по формуле

$$r_b = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{S(x) \cdot S(y)}.$$

Пример 116. Вычислить выборочный коэффициент корреляции по следующим данным:

x_i	2	2	3	5
y_i	4	6	6	8

Решение. Вычислим $\overline{x},\ \overline{x^2},\ \overline{y},\ \overline{y^2},\ \overline{xy},\ S^2(x),\ S^2(y)$:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{4}(2+2+3+5) = 3;$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{4}\left(2^2+2^2+3^2+5^2\right) = 10{,}5;$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{4}(2+6+6+8) = 6;$$

$$\overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 = \frac{1}{4}\left(2^2+6^2+6^2+8^2\right) = 38;$$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i = \frac{1}{4}(8+12+18+40) = 19{,}5;$$

$$S^2(x) = \overline{x^2} - \overline{x}^2 = 10{,}5 - 3^2 = 1{,}5;$$

$$S^2(y) = \overline{y^2} - \overline{y}^2 = 38 - 6^2 = 2,$$

тогда

$$r_b = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{S(x) \cdot S(y)} = \frac{19{,}5 - 18}{\sqrt{3}} \approx 0{,}8666.$$

2.2.20. Метод вычисления выборочного коэффициента корреляции для вариационных рядов

Для вычисления выборочного коэффициента корреляции заполняем корреляционную таблицу. Для этого разбиваем каждый вариационный ряд на интервальный. Затем определяются входящие в формулу для вычисления выборочного коэффициента корреляции параметры.

Пример 117. По данным наблюдений над случайными величинами X и Y получена выборка, приведённая в таблице:

№	X	Y	№	X	Y	№	X	Y
1	7,1	10,0	18	17,2	40,2	35	14,2	21,3
2	9,5	6,7	19	19,9	42,4	36	15,2	25,2
3	11,0	14,0	20	20,1	44,5	37	16,1	21,1
4	12,3	15,1	21	21,7	42,4	38	17,2	24,6
5	11,8	24,2	22	8,5	12,2	39	18,0	23,3
6	14,1	19,9	23	9,7	12,4	40	16,1	30,1
7	15,1	24,3	24	10,2	12,5	41	18,2	27,2
8	14,7	22,2	25	11,1	12,9	42	19,1	30,9
9	16,1	21,0	26	11,3	16,1	43	17,9	35,1
10	13,1	30,1	27	10,9	18,2	44	18,7	36,1
11	13,8	28,1	28	11,4	18,7	45	12,4	17,6
12	16,9	30,3	29	12,3	17,6	46	12,5	18,6
13	19,1	27,3	30	13,2	18,1	47	12,7	19,2
14	14,8	35,3	31	13,1	24,1	48	14,1	26,2
15	17,2	36,3	32	13,6	21,3	49	14,6	27,4
16	19,2	37,4	33	13,7	19,8	50	14,9	30,1
17	22,3	38,0	34	14,6	24,1			

Определить оптимальные длины интервалов и их количество.

Решение. Найдём оптимальные длины интервалов и количество интервалов, используя формулу Стэрджеса:

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \lg n}.$$

Для переменной x

$$h_x = \frac{22{,}3 - 7{,}1}{1 + 3{,}322 \lg 50} = 2{,}2,$$

тогда оптимальное число интервалов равно семи: $[7{,}1;\, 9{,}3)$, $[9{,}3;\, 11{,}5)$, $[11{,}5;\, 13{,}7)$, $[13{,}7;\, 15{,}9)$, $[15{,}9;\, 18{,}1)$, $[18{,}1;\, 20{,}3)$, $[20{,}3;\, 22{,}5)$.

Для переменной y

$$h_y = \frac{44{,}5 - 6{,}7}{1 + 3{,}322 \lg 50} = 6{,}3,$$

тогда оптимальное число интервалов равно шести: $[6{,}7;\, 13{,}0)$, $[13{,}0;\, 19{,}3)$, $[19{,}3;\, 25{,}6)$, $[25{,}6;\, 31{,}9)$, $[31{,}9;\, 38{,}2)$, $[38{,}2;\, 44{,}5)$.

Заполним таблицу согласно рассчитанным интервалам (вместо интервалов в таблице запишем их середины):

Y \ X		1	2	3	4	5	6	7	n_y
		8,2	10,4	12,6	14,8	17,0	19,2	21,4	
1	9,85	2	4						6
2	16,15		4	6					10
3	22,45			3	7	4			14
4	28,75			1	4	2	3		10
5	35,05				1	2	2	1	6
6	41,35					1	1	2	4
n_x		2	8	10	12	9	6	3	50

Для упрощения расчётов перейдём к условным вариантам:

$$u_i = \frac{x_i - C_x}{h_x} = \frac{x_i - 14{,}8}{2{,}2};$$

$$v_i = \frac{y_i - C_y}{h_y} = \frac{y_i - 22{,}45}{6{,}3}.$$

Составим расчётную таблицу:

v	u							n_v	$v_i \cdot n_v$	$v_i^2 \cdot n_v$
	–3	–2	–1	0	1	2	3			
–2	2	4						6	–12	24
–1		4	6					10	–10	10
0			3	7	4			14	0	0
1			1	4	2	3		10	10	10
2				1	2	2		6	12	24
3					1	1	2	4	12	36
n_u	2	8	10	12	9	6	3	50	12	104
$u_i \cdot n_u$	–6	–16	–10	0	9	12	9	–2		
$u_i^2 \cdot n_u$	18	32	10	0	9	24	27	120		
$\overline{v}_i$	–2	–1,5	–0,5	–0,5	1	$\frac{5}{3}$	$\frac{8}{3}$			
$u_i \cdot n_u \cdot \overline{v}_i$	12	24	5	0	9	20	24	94		

Для вычисления выборочного коэффициента корреляции используем формулу

$$r_b = \frac{\overline{uv} - \overline{u} \cdot \overline{v}}{S(u) \cdot S(v)},$$

где

$$\overline{u} = \frac{1}{n} \sum u_i \cdot n_u = \frac{1}{50}(-2) = -0{,}04;$$

$$\overline{u^2} = \frac{1}{n} \sum u_i^2 \cdot n_u = \frac{120}{50} = 2{,}4;$$

$$S(u)=\sqrt{\overline{u^2}-\bar{u}^2}=\sqrt{2{,}4-(-0{,}04)^2}\approx 1{,}549;$$

$$\bar{v}=\frac{1}{n}\sum v_i\cdot n_v=\frac{12}{50}=0{,}24;$$

$$\overline{v^2}=\frac{1}{n}\sum v_i^2\cdot n_v=\frac{104}{50}=2{,}08;$$

$$S(v)=\sqrt{\overline{v^2}-\bar{v}^2}=\sqrt{2{,}08-(0{,}24)^2}\approx 1{,}422;$$

$$\overline{uv}=\frac{1}{n}\sum n_u\cdot u_i\cdot \bar{v}_i=\frac{94}{50}=1{,}88;$$

$$r_b=\frac{\overline{uv}-\bar{u}\cdot\bar{v}}{S(u)\cdot S(v)}=\frac{1{,}88-0{,}04\cdot 0{,}24}{1{,}549\cdot 1{,}422}\approx 0{,}854.$$

2.2.21. Проверка гипотезы о значимости коэффициента корреляции

На практике коэффициент корреляции r обычно неизвестен. По результатам выборки может быть найдена его точечная оценка – выборочный коэффициент корреляции r_b. Равенство нулю выборочного коэффициента корреляции ещё не свидетельствует о равенстве нулю самого коэффициента корреляции, а следовательно, о некоррелированности случайных величин X и Y. Чтобы выяснить, находятся ли случайные величины в корреляционной зависимости, нужно проверить значимость выборочного коэффициента корреляции r_b, то есть установить, достаточна ли его величина для обоснованного вывода о наличии корреляционной связи. Для этого проверяют нулевую гипотезу $H_0: r=0$. Предполагается наличие двухмерного нормального распределения случайных переменных; объём выборки может быть любым. Вычисляют статистику $t=r_b\sqrt{\frac{n-2}{1-r_b^2}}$, которая имеет распределение Стьюдента с $k=n-2$ степенями свободы. Для проверки нулевой гипотезы по уровню значимости α и числу степеней свободы k находят по таблице распределения Стьюдента критическое значение $t_{\alpha;\,k}$. Если $|t| \geq t_{\alpha;\,k}$, то нулевую гипотезу об отсутствии корреляционной связи между переменными X и Y следует отвергнуть. Переменные считают зависимыми. При $|t| < t_{\alpha;\,k}$ нет оснований отвергнуть нулевую гипотезу.

В случае значимого выборочного коэффициента корреляции есть смысл построить доверительный интервал для коэффициента корреляции r. Однако для этого нужно знать закон распределения выборочного коэффициента корреляции r_b. Плотность вероятности выборочного коэффициента корреляции имеет сложный вид, поэтому прибегают к специально подобранным функциям от выборочного коэффициента корреляции, которые сводятся к хорошо изученным распределениям, например к нормальному или Стьюдента. Чаще всего для подбора функции применяют преобразование Фишера. Вычисляют статистику $z = \frac{1}{2}\ln\left(\frac{1+r_b}{1-r_b}\right)$, где $r_b = thz$ – гиперболический тангенс от z.

Распределение статистики z хорошо аппроксимируется нормальным распределением с параметрами $M(z) = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) + \frac{r}{2(n-1)}$, $\sigma_z^2 = \frac{1}{n-3}$. В этом случае доверительный интервал для r имеет вид $thz_1 < r < thz_2$. Величины z_1 и z_2 находятся по формулам

$$z_1 = \frac{1}{2}\ln\left(\frac{1+r_b}{1-r_b}\right) - \frac{z_\alpha}{\sqrt{n-3}};$$

$$z_2 = \frac{1}{2}\ln\left(\frac{1+r_b}{1-r_b}\right) + \frac{z_\alpha}{\sqrt{n-3}},$$

где $\Phi(z_\alpha) = \frac{1-\alpha}{2}$.

Пример 118. Проверить значимость выборочного коэффициента корреляции из примера 53 и найти его доверительный интервал с надёжностью 0,95.

Решение. Для проверки значимости найдём статистику $t = r_b\sqrt{\dfrac{n-2}{1-r_b^2}} = 0{,}854 \cdot \sqrt{\dfrac{48}{1-0{,}854^2}} \approx 11{,}37$. По уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $k = 48$ найдём $t_{\alpha;\,k} = t_{0{,}05;\,47} = 2{,}009$ (см. приложение, табл. 3). Так как $|t| > t_{\alpha;\,k}$, то нулевую гипотезу об отсутствии корреляционной связи между переменными X и Y следует отвергнуть. Следовательно, выборочный коэффициент корреляции значим. Найдём доверительный интервал для выборочного коэффициента корреляции, вычислим:

$$z_1 = \frac{1}{2}\ln\left(\frac{1+0{,}854}{1-0{,}854}\right) - \frac{1{,}96}{\sqrt{47}} \approx 0{,}9849;$$

$$z_2 = \frac{1}{2}\ln\left(\frac{1+0{,}854}{1-0{,}854}\right) + \frac{1{,}96}{\sqrt{47}} \approx 1{,}5566,$$

тогда $thz_1 = th0{,}9849 \approx 0{,}755$ и $thz_2 = th1{,}5566 \approx 0{,}915$. Следовательно, доверительный интервал для выборочного коэффициента корреляции имеет вид $0{,}755 < r < 0{,}915$.

2.2.22. Эмпирическая и теоретическая линии регрессии

Определить форму связи – значит выявить механизм получения зависимой случайной величины.

Кривой регрессии X по Y (или Y по X) называют условное среднее значение случайной величины Y, рассматриваемое как функция определённого класса, параметры которой находятся методом наименьших квадратов по наблюдённым значениям двухмерной случайной величины. То есть уравнение линейной регрессии имеет вид $y = \beta_0 + \beta_1 \cdot x$. Оценке в этом случае подлежат параметры β_0 и β_1, называемые коэффициентами регрессии, а также $\sigma^2_{\text{ост}}$ – остаточная дисперсия. Остаточной дисперсией называется та часть рассеивания результативного признака, которую нельзя объяснить действием наблюдаемого признака. Различают теоретическую и эмпирическую линии регрессии. Теоретическая линия регрессии имеет уравнение

$$y - \overline{y} = r_b \cdot \frac{S_y}{S_x} \cdot \left(x - \overline{x}\right).$$

Эмпирическая линия – это ломаная линия, соединяющая точки $\left(x_i;\, \tilde{y}_i\right)$, где x_i – середина i-го интервала; $\tilde{y}_i$ – расчётное среднее значение для этого интервала.

Пример 119. Построить эмпирическую и теоретическую линии регрессии по данным примера 53.

Решение. Уравнение теоретической линии регрессии имеет вид:

$$y - \overline{y} = r_b \cdot \frac{S_y}{S_x} \cdot \left(x - \overline{x}\right),$$

где

$$\overline{x} = h_x \cdot \overline{u} + C_x = 2{,}2 \cdot \left(-0{,}04\right) + 14{,}8 = 14{,}792;$$

$$\overline{y} = h_y \cdot \overline{v} + C_y = 6{,}3 \cdot 0{,}24 + 22{,}45 = 23{,}962;$$

$$S_x = h_x \cdot S_u = 2{,}2 \cdot 1{,}549 \approx 3{,}408;$$

$$S_y = h_y \cdot S_v = 6{,}3 \cdot 1{,}422 \approx 8{,}959.$$

Тогда уравнение регрессии имеет вид:

$$y - 23{,}962 = 0{,}854 \cdot \frac{8{,}959}{3{,}408} \cdot \left(x - 14{,}792\right)$$

или $y = 2{,}245x - 9{,}245$.

Для построения возьмём середину первого интервала $x = 8{,}2$ и по уравнению регрессии найдём $y = 9{,}2$, затем середину последнего интервала $x = 21{,}4$ и по уравнению регрессии найдём $y \approx 38{,}8$.

Следовательно, для построения эмпирической линии регрессии используем точки с координатами $(8{,}2;\, 9{,}2)$ $(21{,}4;\, 38{,}8)$. При построении эмпирической линии регрессии используем точки вида $(x_i;\, \overline{y}_i)$, где значения $\overline{y}_i$ вычисляем по формуле

$$\overline{y_i} = h_y \cdot \overline{v}_i + C_y.$$

Получаем

$$\overline{y_1} = 6{,}3 \cdot (-2) + 22{,}45 = 9{,}85; \quad \overline{y_2} = 6{,}3 \cdot (-1{,}5) + 22{,}45 = 13;$$

$$\overline{y_3} = 6{,}3 \cdot (-0{,}5) + 22{,}45 = 19{,}3; \quad \overline{y_4} = 6{,}3 \cdot 0{,}5 + 22{,}45 = 25{,}6;$$

$$\overline{y_5} = 6{,}3 \cdot 1 + 22{,}45 = 28{,}75; \quad \overline{y_6} = 6{,}3 \cdot \frac{4}{3} + 22{,}45 = 30{,}85;$$

$$\overline{y_7} = 6{,}3 \cdot \frac{8}{3} + 22{,}45 = 39{,}25.$$

Построим на плоскости точки с координатами: $(8{,}2;\, 9{,}9)$, $(10{,}4;\, 13)$, $(12{,}6;\, 19{,}3)$, $(14{,}8;\, 25{,}6)$, $(17;\, 28{,}8)$, $(19{,}2;\, 30{,}9)$, $(21{,}4;\, 39{,}3)$, соединив их в порядке возрастания x пунктирной линией, получим эмпирическую линию регрессии (Рис. 7).

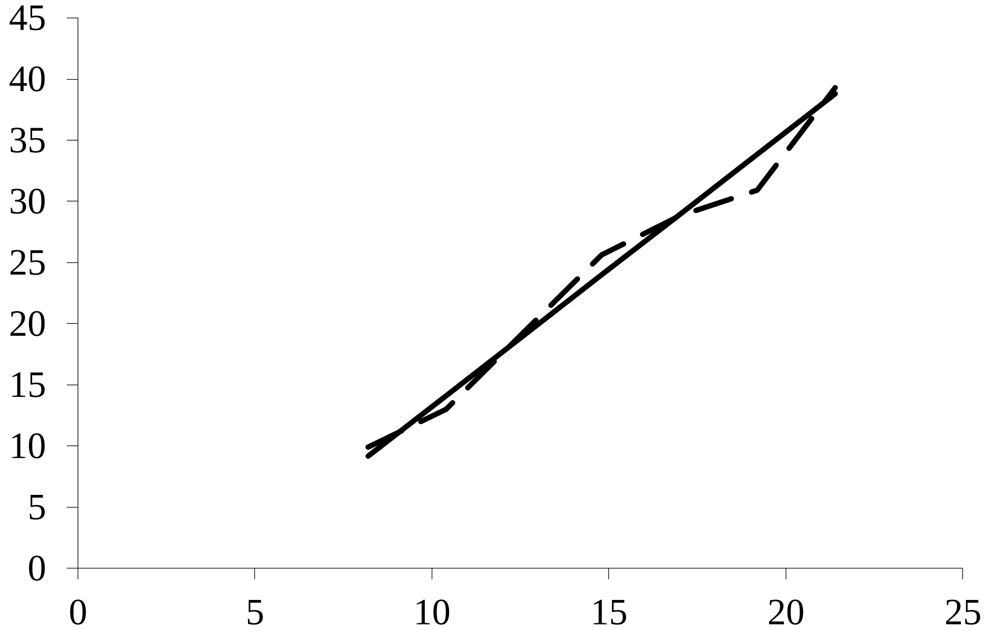

Рис. 7. Эмпирическая линия регрессии.

2.2.23. Значимость коэффициентов регрессии

Проверить значимость оценок коэффициентов регрессии – значит установить, достаточна ли величина оценки для статистически обоснованного вывода о том, что коэффициенты регрессии отличны от нуля. Для этого проверяют гипотезу о равенстве нулю коэффициентов регрессии, соблюдая предпосылки нормальной регрессии.

Для заданной выборки методом наименьших квадратов находим уравнение линии регрессии. Для этого вычисляем статистику $t = \left|\frac{b}{S_b}\right|$, которая имеет $k = n - 2$ степеней свободы, b – оценка коэффициента регрессии, S_b – оценка среднего квадратического отклонения коэффициента, иначе стандартная ошибка оценки. По уровню значимости и числу степеней свободы находим $t_{\alpha;\,k}$. Если $t > t_{\alpha;\,k}$, то гипотезу о равенстве нулю коэффициента регрессии отвергают, следовательно, при заданном уровне значимости коэффициент регрессии значим. Оценки среднего квадратического отклонения находим по формулам

$$S_{b_0} = \frac{S_{\text{ост}}}{\sqrt{n-2}};\ S_{b_1} = \frac{S_{\text{ост}}}{S_x \cdot \sqrt{n-2}},$$

где $S_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$ и $S_{\text{ост}}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - y(x_i)\right)^2$.

Доверительные интервалы для коэффициентов регрессии находятся по формулам $b - t_{\alpha;\,k} \cdot S_b < \beta < b + t_{\alpha;\,k} \cdot S_b$.

Пример 120. Проверить значимость коэффициентов регрессии и найти доверительные интервалы для них при уровне значимости $\alpha = 0{,}05$ по данным предыдущей задачи.

Решение. Уравнение регрессии имеет вид $y = 2{,}245x - 9{,}245$. Найдём остаточную дисперсию

$$S_{\text{ост}}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(\overline{y}_i - y(x_i)\right)^2 \cdot m_x,$$

для этого составим таблицу:

x_i	m_x	$\overline{y}_i$	$y(x_i)$	$\left(\overline{y}_i - y(x_i)\right)^2 \cdot m_x$
8,2	2	9,85	9,164	0,941192
10,4	8	13	14,1	9,732872
12,6	10	19,3	19,04	0,66564
14,8	12	25,6	23,98	31,453932
17,0	9	28,8	28,92	0,1296
19,2	6	30,9	33,86	52,534086
21,4	3	39,3	38,8	0,756012
				96,213334

Остаточная дисперсия равна

$$S_{\text{ост}}^2 = \frac{96{,}213334}{48} \approx 2{,}004,$$

тогда

$$S_{b_0} = \frac{S_{\text{ост}}}{\sqrt{n-2}} \approx 0{,}204;$$

$$S_{b_1} = \frac{S_{\text{ост}}}{S_x \cdot \sqrt{n-2}} \approx 0{,}05996;$$

$$t_{b_0} = \left|\frac{\beta_0}{S_{b_0}}\right| = \left|\frac{-9{,}245}{0{,}204}\right| \approx 45{,}24;$$

$$t_{b_1} = \left|\frac{\beta_1}{S_{b_1}}\right| = \left|\frac{2{,}245}{0{,}05996}\right| \approx 37{,}44.$$

По таблице (см. приложение, табл. 5) находим $t_{\text{кр}}\left(48;\,0{,}05\right)$. Так как $t_{b_0} > t_{\text{кр}}$ и $t_{b_1} > t_{\text{кр}}$, то оба коэффициента значимы. Доверительный интервал для β_0 имеет вид:

$$-9{,}245 - 0{,}204 \cdot 2{,}009 < \beta_0 < -9{,}245 + 0{,}204 \cdot 2{,}009$$

или

$$-9{,}655 < \beta_0 < -8{,}835.$$

Доверительный интервал для β_1 имеет вид:

$$2{,}245 - 0{,}05996 \cdot 2{,}009 < \beta_1 < 2{,}245 + 0{,}05996 \cdot 2{,}009$$

или

$$2{,}1245 < \beta_1 < 2{,}3655.$$

Пример 121. Найти оценки коэффициентов уравнения регрессии $y = \beta_0 + \beta_1 \cdot x$, проверить их значимость и построить доверительные интервалы при уровне значимости $\alpha = 0{,}05$ по данной выборке:

x_i	1	2	3	4	5	6	7	8	9	10
y_i	7	9	11	15	15	19	20	23	24	27

Решение. Найдём оценки коэффициентов регрессии по формулам

$$b = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} \text{ и } a = \overline{y} - \beta_1 \cdot \overline{x}.$$

Вычислим значения входящих в формулы величин:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}(1+2+3+4+5+6+7+8+9+10) = 5{,}5;$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{10}(7+9+11+15+15+19+20+23+24+27) = 17;$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 =$$

$$= \frac{1}{10}\left(1^2+2^2+3^2+4^2+5^2+6^2+7^2+8^2+9^2+10^2\right) = 38{,}5;$$

$$\overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 =$$

$$= \frac{1}{10}\left(7^2+9^2+11^2+15^2+15^2+19^2+20^2+23^2+24^2+27^2\right) = 329{,}6;$$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i =$$

$$= \frac{1}{10}(7 + 18 + 33 + 60 + 75 + 114 + 140 + 184 + 116 + 270) =$$

$$= 111{,}1.$$

Тогда

$$b = \frac{111{,}1 - 5{,}5 \cdot 17}{38{,}5 - 5{,}5^2} \approx 2{,}13;$$

$$a = 17 - 2{,}13 \cdot 5{,}5 \approx 5{,}27.$$

Следовательно, уравнение регрессии имеет вид $y = 2{,}13x + 5{,}27$.

Проверим значимость коэффициентов регрессии. Для этого вычислим:

$$S_x = \sqrt{\overline{x^2} - \overline{x}^2} = \sqrt{38{,}5 - 5{,}5^2} \approx 2{,}872.$$

Найдём $y(x_i)$, используя уравнение регрессии:

$$y(1) \approx 7{,}4, \quad y(2) \approx 9{,}53, \quad y(3) \approx 11{,}66;$$

$$y(4) \approx 13{,}79, \quad y(5) \approx 15{,}92, \quad y(6) \approx 18{,}05;$$

$$y(7) \approx 20{,}18, \quad y(8) \approx 22{,}31, \quad y(9) \approx 24{,}44, \quad y(10) \approx 26{,}57.$$

Найдём остаточную дисперсию по формуле

$$S^2_{\text{ост}} = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - y(x_i))^2$$

$$= \frac{1}{8}\Big((7 - 7{,}4)^2 + (9 - 9{,}53)^2 + (11 - 11{,}66)^2 + (15 - 13{,}79)^2 +$$

$$+ (15 - 15{,}92)^2 + (19 - 18{,}05)^2 + (20 - 20{,}18)^2 +$$

$$+ (23 - 22{,}31)^2 + (24 - 24{,}44)^2 + (27 - 26{,}57)^2\Big) \approx 0{,}622$$

Тогда

$$S_{\text{ост}} = \sqrt{0{,}622} \approx 0{,}7887;$$

$$S_{b_0} = \frac{S_{\text{ост}}}{\sqrt{n-2}} = \frac{0{,}7887}{\sqrt{8}} \approx 0{,}279;$$

$$S_{b_1} = \frac{S_{\text{ост}}}{S_x \cdot \sqrt{n-2}} = \frac{0{,}7887}{2{,}872 \cdot \sqrt{8}} \approx 0{,}097.$$

Статистики равны:

$$t_0 = \left| \frac{a}{S_a} \right| = \frac{5{,}27}{0{,}279} \approx 18{,}9;$$

$$t_1 = \left| \frac{b}{S_b} \right| = \frac{2{,}13}{0{,}097} \approx 21{,}96.$$

По таблице (см. приложение, табл. 5) находим $t_{\text{кр}}(8;\, 0{,}05) = 2{,}31$. Так как $t_0 > t_{\text{кр}}$ и $t_1 > t_{\text{кр}}$, то оба коэффициента значимы. Доверительный интервал для β_0 имеет вид:

$$5{,}27 - 0{,}279 \cdot 2{,}31 < \beta_0 < 5{,}27 + 0{,}279 \cdot 2{,}31$$

или

$$4{,}625 < \beta_0 < 5{,}914.$$

Доверительный интервал для β_1 имеет вид:

$$2{,}13 - 0{,}097 \cdot 2{,}31 < \beta_1 < 2{,}13 + 0{,}097 \cdot 2{,}31$$

или

$$1{,}906 < \beta_1 < 2{,}354.$$

2.2.24. Корреляционное отношение

На практике часто предпосылки корреляционного анализа нарушаются: один из признаков оказывается неслучайным или признаки не имеют нормального распределения. Для изучения связи между ними в этом случае существует показатель зависимости признаков, основанный на показателе изменчивости общей (или полной) дисперсии.

Полной называется дисперсия признака относительно его математического ожидания. Так для признака Y это $\sigma_y^2 = M(y - M(y))^2$.

Дисперсию σ_y^2 можно разложить на две составляющие, одна из которых характеризует влияние фактора X на Y, другая – влияние прочих факторов. Очевидно, чем меньше влияние прочих факторов, тем теснее связь, тем более приближается она к функциональной.

По выборочным данным рассчитываем корреляционное отношение:

$$\eta^2 = \frac{S^2_{y/x}}{S_y^2},$$

где

$$S_y^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 \cdot m_i \quad \text{и} \quad S^2_{y/x} = \frac{1}{n}\sum_{i=1}^{n}\left(y(x_i) - \overline{y}\right)^2 \cdot m_i .$$

Значения η^2, лежащие в интервале $0 \le \eta^2 \le 1$, являются показателями тесноты группировки точек около кривой регрессии независимо от её вида (формы связи). Если зависимость является линейной, то корреляционное отношение совпадает с коэффициентом корреляции:

$$\eta^2 = \frac{S^2_{y/x}}{S_y^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y(x_i) - \overline{y}\right)^2 \cdot m_i}{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 \cdot m_i} = \frac{\sum_{i=1}^{n}\left(a + b\cdot x_i - a - b\cdot \overline{x}\right)^2}{S_y^2} =$$

$$= \frac{b \cdot S_x^2}{S_y^2} = \frac{r^2 \cdot \frac{S_y^2}{S_x^2} \cdot S_x^2}{S_y^2} = r^2.$$

Если зависимость является нелинейной, то $0 \le r^2 \le \eta^2$.

Пример 122. Вычислить корреляционное отношение по выборке:

x_i	2	4	6	8	10	12	14	16	18	20
y_i	10	4	0	1	4	10	20	32	50	69

Предполагая, что зависимость между переменными имеет вид: $y = ax^2 + bx + c$.

Решение. Согласно методу наименьших квадратов для определения коэффициентов имеем систему уравнений:

$$\begin{cases} a\sum\limits_{i=1}^{n} x_i^2 + b\sum\limits_{i=1}^{n} x_i + c\cdot n = \sum\limits_{i=1}^{n} y_i; \\ a\sum\limits_{i=1}^{n} x_i^3 + b\sum\limits_{i=1}^{n} x_i^2 + c\cdot \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i\cdot y_i; \\ a\sum\limits_{i=1}^{n} x_i^4 + b\sum\limits_{i=1}^{n} x_i^3 + c\cdot \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i^2\cdot y_i. \end{cases}$$

Составим таблицу для расчёта параметров зависимости:

№	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i \cdot y_i$	$x_i^2 \cdot y_i$	y_i^2
1	2	10	4	8	16	20	40	100
2	4	4	16	64	256	16	64	16
3	6	0	36	216	1296	0	0	0
4	8	1	64	512	4096	8	64	1
5	10	4	100	1000	10000	40	400	16
6	12	10	144	1728	20736	120	1440	100
7	14	20	196	2744	38416	280	3920	400
8	16	32	256	4096	65536	512	8192	1024
9	18	50	324	5832	104976	900	16200	2500
10	20	69	400	8000	160000	1380	27600	4761
$\sum$	110	200	1540	24200	405328	3276	57920	8918

Получаем систему уравнений:

$$\begin{cases} 1540a + 110b + 10c = 200; \\ 24200a + 1540b + 110c = 3275; \\ 405328a + 24200b + 1540c = 57920. \end{cases}$$

Следовательно, зависимость имеет вид:

$$y = 0{,}4081x^2 - 5{,}7186x + 20{,}05.$$

Найдём значение корреляционного отношения, для этого составим таблицу:

№	x_i	y_i	$y(x_i)$	$(y_i - \bar{y})^2$	$(y(x_i) - \bar{y})^2$
1	2	10	10,24546	100	95,15113
2	4	4	3,706064	256	265,4924
3	6	0	0,431824	400	382,9135
4	8	1	0,422736	361	383,2693
5	10	4	3,6788	256	266,3816
6	12	10	10,20002	100	96,03969
7	14	20	19,98638	0	0,000185
8	16	32	33,0379	144	169,9869
9	18	50	49,35458	900	861,6911
10	20	69	68,9364	2401	2394,771
$\sum$	110	200	200,0002	4918	4915,697

Тогда корреляционное отношение равно

$$\eta^2 = \frac{4915{,}697}{4918} \approx 0{,}999532.$$

Коэффициент корреляции равен $r^2 = 0{,}7134.$

Задачи для самостоятельной решение

2.1. При исследовании количественного признака X из генеральной совокупности была получена выборка

5,7,4,6,5,5,5,7,5,6,6,5,5,6,4,6,5,6,4,5.

Найти объём выборки, построить вариационный ряд выборки и её статистическое распределение.

Вариационный ряд выборки: 4,4,4,5,5,5,5,5,5,5,5,5,6,6,6,6,6,6,7,7. Статическое распределение:

x_i	4	5	6	7
n_i	3	9	6	2

2.2. Выборка задана распределением частот:

x_i	-3	-1	0	4	8	13	17
n_i	1	3	4	7	5	8	12

Найти распределением относительных частот.

2.3. Выборка задана интервальным распределением частот:

x_i	(0:2]	(2:4]	(4:6]	(6:8]	(8:10]
n_i	5	14	16	11	4

Найти распределением относительных частот.

2.4. Искомый полигон частот для заданного распределение выборки:

x_i	3	5	8	14	16	20	22
n_i	32	41	18	56	23	12	47

2.5. Задано интервальное статистическое распределение выборки:

$(x_i; x_{i+1}]$	(1;3]	(3;5]	(5;7]	(7;9]	(9;11]	(11;13]	(13;15]
ω_i	0.15	0.05	0.2	0.1	0.05	0.25	0.2

Найти выборочное среднее, выборочную дисперсию и среднее квадратическое отклонение выборки.

2.6. На одном из отрезков автострады планируется обустроить остановку автобуса. Распределение населенных пунктов с численностью их населения в таблице.

На каком километре автострады расположен населённый пункт, км	3	6	8	14	20	22	25
Численность населения, тыс. чел.	4	3	5	6	2	3	1

На каком километре автострады нужно расположить эту остановку, чтобы суммарное расстояние, которое будут покрывать потенциальные пассажир до этой остановки, было наименьшим.

2.7. Задано интервальное статистическое распределения выборки:

$(x_i;x_{i+1}]$	(4;8]	(8;12]	(12;16]	(16;20]	(20;24]	(24;28]	(28;32]
n_i	3	6	6	10	14	5	1

Найти медиану, моду и вариационный размах.

2.8. Найти методом произведений выборочное среднее и выборочную дисперсию для такого распределения выборки объёма $n=100$:

x_i	40	45	50	55	60	65	70
n_i	12	18	20	24	14	7	5

Указание. В качестве ложного нуля выбрать значение четвёртой варианты: $C=x_4=55$.

2.9. Найти методом произведений выборочное среднее и выборочную дисперсию для такого распределения выборки объёма $n=100$.

x_i	5	7	10	11	14	15	18	20	22	25
n_i	6	8	11	18	22	11	7	7	5	5

Указание. Разбить весь интервал на пять частных интервалов, в качестве ложного нуля выбрать значение середин третьего интервала. При определение выборочной дисперсии учесть поправку Шепарда.

2.10. Найти методом сумм выборочное среднее и выборочную дисперсию для заданного статистического распределения выборки объема $n=100$.

x_i	42	49	56	63	70	77	84	91	98	105
n_i	2	5	8	19	26	14	10	8	5	3

Указание. В качестве ложного нуля выбрать значение пятой варианты: $C=x_5=70$.

2.11. Найти методом наибольшего правдоподобия оценку параметра p биномиального распределения, если в n_1 независимых испытаниях событие A появилось m_1 раз и в n_2 независимых испытаниях событие появилось m_2 раз.

2.12. Используя метод наибольшего правдоподобия оценку параметры a и σ^2 нормального распределения, если в результате n независимых испытаний случайная величина ξ приняла значения $\xi_1, \xi_2, \xi_3, ..., \xi_n$.

2.13. Случайная величина X (число появлений события A в m независимых испытаниях) подчинена закону распределения Пуассона с неизвестным параметром λ. Найти методом наибольшего правдоподобия по выборке $x_1, x_2, ..., x_n$ точечную оценку неизвестного параметра λ распределения Пуассона.

2.14. Случайная величина-время безотказной работы изделия имеет показательное распределение. В таблице приведены данные по времени работ в часах для 1000 изделий. Найти методом максимального правдоподобия точечную оценку неизвестного параметра λ.

2.15. Найти методом наибольшего правдоподобия по выборке $x_1, x_2, ..., x_n$ точечную оценку параметра p геометрического распределения:

$$P(X = x_i) = (1-p)^{x_i - 1} \cdot p,$$

где x_1-число испытаний, произведенных до появления события, p-вероятность появления события в одном испытании.

2.16. Методом максимального правдоподобия найти точечную оценку параметра λ по данной выборке.

x	1-3	3-5	5-7	7-9	9-11	11-13	13-15	15-17	17-19
n	5	6	7	15	22	30	34	35	

при условии, что соответствующая непрерывная случайная величина имеет плотность распределения $f(x) = \lambda \exp(\lambda(x-20)), x \le 20$.

2.17. Методом наименьших квадратов для данных, представленных в таблице, найти линейную зависимость $y = ax + b$. Данные

i	1	2	3	4	5	6	7	8
x_i	-2	-1	0	1	2	3	4	5
y_i	-11,47	-7,59	-4,32	-0,41	3,01	6,91	10,12	14,08

2.18. Данные наблюдений над случайной двумерной величиной (*X,Y*) представлены в корреляционной таблице. Методом наименьших квадратов найти выборочное уравнение прямой регрессии *Y* на *X*.

X	*Y*					n_x
	10	20	30	40	50	
3	7	-	-	-	-	7
8	11	5	-	-	-	16
13	-	19	15	5	-	39
18	-	3	15	6	1	25
23	-	-	2	4	4	10
28	-	-	-	-	3	3
n_y	18	27	32	15	8	100

2.19. Считая, что зависимость между переменным *x* и *y* имеет вид $y = ax^2 + bx + c,$ найти оценки параметров a, b и c методом наименьших квадратов по выборке:

X	7	31	61	99	129	178	209
Y	13	10	9	10	12	20	26

2.20. Проводится анализ взаимосвязи количества населения практикующих врачей (*Y*) в регионе.

Годы	81	82	83	84	85	86	87	88	89	90
X, млн. чел.	10	10,3	10,4	10,55	10,6	10,7	10,75	10,9	10,9	11
Y, тыс. чел.	12,1	12,6	13	13,8	14,9	16	18	20	21	22

Оцените по МНК коэффициенты линейного уравнения регрессии $y = b_0 + b_1 x$. Существенно ли отличаются от нуля найденные коэффициенты?

Проверьте значимость полученного уравнения при $\alpha = 0{,}01$.

Если количество населения в 1995 году составит 11,5 млн. чел., каково ожидаемое количество врачей? Рассчитайте 99%-й доверительный интервал для данного прогноза. Рассчитайте коэффициент детерминации.

2.21. Используя критерий Пирсона, при уровне значимости 0,05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности по результатам выборки:

X:	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3
N:	7	9	28	27	30	26	21	25	22	9	5

2.22. Были исследованы 200 готовых деталей на отклонения истинного размера от расчетного. Сгруппированные данные приведены в следующей таблице:

По данному статистическому ряду построить гистограмму. По виду гистограммы выдвинуть гипотезу о виде закона распределения (например, предположить, что исследуемая величина имеет нормальный закон распределения). Подобрать параметры закона распределения (равные их оценкам на основе опытных данных). На том же графике построить функцию плотности вероятности, соответствующую выдвинутой гипотезе. С помощью критерия согласия проверить, согласуется ли гипотеза с опытными данными. Уровень значимости взять, например, равным 0,05.

Критерий Пирсона, распределение по закону Пуассона

2.23. Отдел технического контроля проверил n партий однотипных изделий и установил, что число нестандартных изделий в одной партии имеет эмпирическое распределение, приведенное в таблице, в одной строке которой указано количество x_i нестандартных изделий в одной партии, а в другой строке – количество n_i партий, содержащих x_i нестандартных изделий. Требуется при уровне значимости, а 0,05 проверить гипотезу о том, что случайная величина X (число нестандартных изделий в одной партии) распределена по закону Пуассона.

2.24. В результате обследования 150 человек были получены данные о количестве приобретаемых за месяц цветных иллюстрированных журналов. Соответствует ли данное распределение закону редких событий Пуассона? Критерий Пирсона, распределение по показательному закону.

2.25. В итоге испытаний 1000 элементов в таблице. Требуется при уровне значимости проверить гипотезу о том, что данные в генеральной совокупности распределены по показательному закону.

Время безотказной работы 0-10 10-20 20-30 30-40 40-50 50-60 60-70

Число отказавших элементов 365 245 150 100 70 45 25

Критерий Пирсона, распределение по равномерному закону.

2.26. В некоторой местности в течение 300 суток регистрировалась среднесуточная температура воздуха. В итоге наблюдений было получено эмпирическое распределение, приведенное в таблице 40 (в первом столбце указан интервал температуры в градусах, во втором столбце – частота n_i, т.е. количество дней, среднесуточная температура которых принадлежит этому интервалу). Требуется при уровне значимости 0,05 проверить гипотезу о том, что среднесуточная температура воздуха распределена равномерно.

2.27. По данным 7 измерений некоторой величины найдены средняя результатов измерений, равная 30 и выборочная дисперсия, равная 36. Найдите границы, в которых с надежностью 0,99 заключено истинное значение измеряемой величины.

2.28. Строительная компания хочет оценить среднюю стоимость ремонтных работ, выполняемых для клиентов. Каким должен быть объем выборки среди 1200 клиентов строительной фирмы, если среднее квадратическое отклонение по результатам пробного обследования составило 850 у.е., а предельная ошибка выборки не должна превышать 200 у.е. с вероятностью 0,95?

2.29. Из партии объемом 500 однородных товаров для проверки по схеме случайной бесповторной выборки отобрано 70 товаров, среди которых оказалось 56 набракованных. Найдите вероятность того, что доля бракованных товаров во всей партии отличается от полученной доли в выборке не более чем на 0,02 (по абсолютной величине), а также границы, в которых с надежностью 0,96 заключено доля бракованных товаров во всей партии.

2.30. По данным выборочного контроля найти выборочные математическое ожидание и дисперсию нормальной случайной ξ. Найти доверительные интервалы для них, соответствующие доверительной вероятности 0,9.

2.31. С целью размещения рекламы опрошено 420 телезрителей, из которых данную передачу смотрят 170 человек. С доверительной вероятностью $\gamma = 0,91$ найти долю телезрителей, охваченных рекламой в лучшем случае.

2.32. Найти доверительный интервал с надёжностью 0,999 для оценки неизвестного математического ожидания a нормально распределённой случайной величины X, если дисперсия этой случайной величины $\sigma^2 = 25$, выборочное среднее $\overline{x}_n = 3$, а объём выборки $n = 36$.

2.33. На одном и том же станке изготовляются детали. Каким должна быть минимальное количество деталей для того, чтобы с надёжностью 0,95 точность оценки математического ожидания a случайной величины X, которая характеризует длину детали, по выборочному среднему составляла $\varepsilon = 3\text{мм}$, если известно среднее квадратическое отклонение случайной величины $X : \sigma = 5\text{мм}$?

Считается, что случайная величина X имеет нормальный закон распределения.

2.34. В результате статистических исследований случайной величины X получена выборка объёма $n = 50$ с таким статистическим распределением:

$$\frac{x_i}{n_i}\left|\frac{-3}{2}\right|\frac{-2}{4}\left|\frac{-1}{8}\right|\frac{0}{11}\left|\frac{1}{11}\right|\frac{2}{9}\left|\frac{3}{5}\right. .$$

Найти с надёжностью $\alpha = 0{,}999$ интервальную оценку математического ожидания a случайной величины X по выборочному среднему.

Считается, что случайная величина X нормально распределена.

2.35. По данным выборки объёма $n = 25$ найдено “исправленное” выборочное среднее квадратическое отклонении $s = 1$ нормальна распределённой случайной величине X.

Найти доверительный интервал для среднего квадратического отклонения σ случайной величины X с надёжностью $\alpha = 0{,}999$.

2.36. В результате статических исследований случайной величины X получена выборка объёма $n = 10$ с таким статистическим распределением:

$$\frac{x_i}{n_i}\left|\frac{-1}{1}\right|\frac{0}{3}\left|\frac{1}{4}\right|\frac{2}{2}.$$

Найти с надёжностью $\alpha = 0{,}99$ интервальную оценку среднего квадратического отклонения σ случайной величины X.

Считается, что случайная величина X распределена по нормальному закону.

2.37. Проводятся независимые с одинаковой, но неизвестной вероятностью p успеха. Найти доверительный интервал для оценки вероятности p с надёжностью $\alpha = 0{,}95$, если из 100 испытаний успешными были 27.

2.38. Во время испытания стали на прочность оказалось, что из 1000 стальных прутьев не выдержали испытания 57.

Найти с надёжностью $\alpha = 0{,}999$ интервальную оценку вероятности того, что стальной прут не пройдёт испытания на прочность.

2.39. Имеются данные средней выработки на одного рабочего Y (тыс. руб.) и товар оборота X (тыс. руб.) в 20 магазинах за квартал. На основе указанных данных требуется:

1) определить зависимость (коэффициент корреляции) средней выработки на одного рабочего от товарооборота,

2) составить уравнение прямой регрессии этой зависимости.

2.40. С целью анализа взаимного влияния зарплаты и текучести рабочей силы на пяти однотипных фирмах с одинаковым числом работников проведены измерения уровни месячной зарплаты X и числа уволившихся за год рабочих Y:

X : 100 150 200 250 300

Y : 60 35 20 20 15

Найти линейную регрессию Y на X, выборочный коэффициент корреляции.

2.41. Найти выборочные числовые характеристики и выборочное уравнение линейной регрессии $y_x = ax + b$. Построить прямую регрессии и изобразить на плоскости точки (x, y) из таблицы. Вычислить остаточную дисперсию. Проверить адекватность линейной регрессионной модели по коэффициенту детерминации.

2.42. Вычислить коэффициенты уравнения регрессии. Определить выборочный коэффициент корреляции между плотностью древесины маньчжурского ясеня и его прочностью. Решая задачу необходимо построить поле корреляции, по виду поля определить вид зависимости, написать общий вид уравнения регрессии Y на X, определить коэффициенты уравнения регрессии и вычислить коэффициенты корреляции между двумя заданными величинами.

2.43. Компанию по прокату автомобилей интересует зависимость между пробегам автомобилей X и стоимостью ежемесячного технического обслуживания Y. Для выяснения характера этой связи было отобрано 15 автомобилей. Постройте график исходных данных и определите по нему характер зависимости. Рассчитайте выборочный коэффициент линейной корреляции Пирсона, проверьте его значимость при 0,05. Постройте уравнение регрессии и дайте интерпретацию полученных результатов.

2.44. На основании 18 наблюдений установлено, что на 64% вес (X) кондитерских изделий зависит от их объёма (Y). Можно ли на уровне значимости $\alpha = 0,05$ утверждать, что между X и Y существует зависимости?

2.45. Исследование 27 семей по среднедушевому доходу (X) и сбережениям (Y) дало результаты: $\overline{X} = 82$ у.е., $S_x = 31$ у.е., $\overline{Y} = 39$ у.е., $S_y = 29$ у.е., $\overline{XY} = 3709$ у.е. 2. При $\alpha = 0,05$ проверить наличие линейной связи между X и Y. Определить размер сбережений семей, имеющих среднедушевой доход $X = 130$ у.е.

Вопросы для повторения второй главы.

2.1. Что такое выборочная совокупность (выборка), генеральная совокупность, объем совокупности?

2.2. Дать определение повторной, бесповторной и репрезентативной выборкам.

2.3. Дать определение простого случайного и типического отборов.

2.4. Что представляет собой механический и серийный отборы?

2.5. Дать определение вариантам, вариационным рядам, частотам и относительным частотам.

2.6. Статистическое распределение выборки и как оно задается, какова разница между распределением в теории вероятностей и распределением в математической статистике.

2.7. Эмпирическая функция распределения и теоретическая функция распределения.

2.8. Какими свойствами обладает эмпирическая функция распределения ?

2.9. Что называется полигоном частот и полигоном относительных частот, как они строятся?

2.10. Гистограмма частот, как она строится и чему равна площадь гистограммы частот?

2.11. Гистограмма относительных частот, как она строится и чему равна площадь гистограммы относительных частот.

2.12. Что называется статистической оценкой неизвестного параметра, и какими важнейшими свойствами она может обладать?
2.13. Несмещенная оценка и чем обосновывается ее введение?
2.14. Эффективная оценка и в чем необходимость ее ввода?
2.15. Смещенная и состоятельная оценки?
2.16. Генеральная средняя и по каким формулам она вычисляется?
2.17. Что называется выборочной средней и по каким формулам она вычисляется?
2.18. Какой оценкой генеральной средней является выборочная средняя?
2.19. Генеральная дисперсия и по каким формулам она вычисляется?
2.20. Что называется выборочной дисперсией, и по каким формулам она вычисляется?
2.21. Генеральное среднее квадратическое отклонение и выборочное среднее квадратическое отклонение.
2.22. Дать определение несмещенной оценки генеральной дисперсии.
2.23. Точность оценки и надежность (доверительная вероятность)?
2.24. Что называется доверительным интервалом?
2.25. Как находится доверительный интервал для оценки математического ожидания нормального распределения при известном среднем квадратическом отклонении?
2.26. Что вы знаете о распределениях «хи квадрат» и Стьюдента?
2.27. Как находится доверительный интервал для оценки математического ожидания нормального распределения при неизвестном среднем квадратическом отклонении?
2.28. Как находится доверительный интервал для оценки среднего квадратического отклонения нормального распределения?
2.29. Что изучают корреляционный и регрессионный анализ, как могут быть связаны случайные величины, что такое функция случайного аргумента и функциональная зависимость?

2.30. Что такое условное среднее, выборочное уравнение регрессии, выборочная регрессия, выборочная линия регрессии, и какие две задачи теории корреляции вы знаете?

2.31. В каком виде ищется выборочное уравнение прямой линии регрессии по не сгруппированным данным и что такое выборочный коэффициент регрессии?

2.32. В чем суть метода наименьших квадратов и как с его помощью находится выборочное уравнение прямой линии регрессии?

2.33. Что вы знаете о корреляционной таблице?

2.34. Как находятся параметры выборочного уравнения прямой линии регрессии по сгруппированным данным?

2.35. Дать определение корреляционным моментам и коэффициентам корреляции?

2.36. Коррелированные и некоррелированные случайные величины, и какова связь между понятиями зависимости и коррелированности случайных величин?

2.37. Дать определение выборочным коэффициентам корреляции?

2.38. Выборочное корреляционное отношение и для чего оно служит?

2.39. Какие свойства выборочного корреляционного отношения вы знаете?

2.40. Что вы понимаете под статистической гипотезой? Приведите примеры.

2.41. Нулевая (основная), конкурирующая (альтернативная), простая, сложная гипотезы?

2.42. Что называется наблюдаемым значением критерия, критической областью, областью принятия гипотезы (областью допустимых значений)?

2.43. Критические точки (границы), правосторонняя, левосторонняя, односторонняя, двусторонняя критическая области?

2.44. Уровень значимости и как находится критическая область?

2.45. Мощность критерия и как она связана с ошибкой второго рода ?

2.46. Критерий согласия и как применяется критерий Пирсона?

2.47. По каким причинам различаются эмпирические и теоретические частоты?

2.48. Какая случайная величина принимается в качестве критерия проверки нулевой гипотезы о нормальном распределении генеральной совокупности, и ее свойства?

2.49. В чем суть правила проверки нулевой гипотезы о нормальном распределении генеральной совокупности?

2.50. Каким способом находятся теоретические частоты?

Приложение

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

x	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	0,3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	0,3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	0,3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	0,3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	0,3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	0,3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	0,3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	0,2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	0,2261	2637	2613	2589	2565	2541	256	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	0,2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	0,1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	0,1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	0,1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	0,1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	0,1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0,0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0,0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0,0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0,0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0,0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0,0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0,0224	0219	0213	0208	0203	0198	0194	0189	0184	0180

2,5	0,0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0,0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0,0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0,0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0,0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0,0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0,0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0,0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0,0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0,0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0,0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0,0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0,0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0,0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Функция $\varphi(x)$ является чётной, $\varphi(x) \approx 0$ при $x > 4$.

Таблица 2. Значения функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,45	0,1736	0,90	0,3159	1,35	0,4115	1,80	0,4641
0,01	0,0040	0,46	0,1772	0,91	0,3186	1,36	0,4131	1,81	0,4649
0,02	0,0080	0,47	0,1808	0,92	0,3212	1,37	0,4147	1,82	0,4656
0,03	0,0120	0,48	0,1844	0,93	0,3238	1,38	0,4162	1,83	0,4664
0,04	0,0160	0,49	0,1879	0,94	0,3264	1,39	0,4177	1,84	0,4671
0,05	0,0199	0,50	0,1915	0,95	0,3289	1,40	0,4192	1,85	0,4678
0,06	0,0239	0,51	0,1950	0,96	0,3315	1,41	0,4207	1,86	0,4686
0,07	0,0279	0,52	0,1985	0,97	0,3340	1,42	0,4222	1,87	0,4693
0,08	0,0319	0,53	0,2019	0,98	0,3365	1,43	0,4236	1,88	0,4699

0,09	0,0359	0,54	0,2054	0,99	0,3389	1,44	0,4251	1,89	0,4706
0,10	0,0398	0,55	0,2088	1,00	0,3413	1,45	0,4265	1,90	0,4713
0,11	0,0438	0,56	0,2123	1,01	0,3438	1,46	0,4279	1,91	0,4719
0,12	0,0478	0,57	0,2157	1,02	0,3461	1,47	0,4292	1,92	0,4726
0,13	0,0517	0,58	0,2190	1,03	0,3485	1,48	0,4306	1,93	0,4732
0,14	0,0557	0,59	0,2224	1,04	0,3508	1,49	0,4306	1,94	0,4738
0,15	0,0596	0,60	0,2257	1,05	0,3531	1,50	0,4332	1,95	0,4744
0,16	0,0636	0,61	0,2291	1,06	0,3554	1,51	0,4345	1,96	0,4750
0,17	0,0675	0,62	0,2324	1,07	0,3577	1,52	0,4357	1,97	0,4756
0,18	0,0714	0,63	0,2357	1,08	0,3599	1,53	0,4370	1,98	0,4761
0,19	0,753	0,64	0,2389	1,09	0,3621	1,54	0,4382	1,99	0,4767
0,20	0,0793	0,65	0,2422	1,10	0,3643	1,55	0,4394	2,00	0,4772
0,21	0,0832	0,66	0,2452	1,11	0,3665	1,56	0,4406	2,02	0,4783
0,22	0,0871	0,67	0,2486	1,12	0,3686	1,57	0,4418	2,04	0,4793
0,23	0,0910	0,68	0,2517	1,13	0,3708	1,58	0,4429	2,06	0,4803
0,24	0,0948	0,69	0,2549	1,14	0,3729	1,59	0,4441	2,08	0,4812
0,25	0,0987	0,70	0,2580	1,15	0,3749	1,60	0,4452	2,10	0,4821
0,26	0,1026	0,71	0,2611	1,16	0,3770	1,61	0,4463	2,12	0,4830
0,27	0,1064	0,72	0,2642	1,17	0,3790	1,62	0,4474	2,14	0,4838
0,28	0,1103	0,73	0,2673	1,18	0,3810	1,63	0,4484	2,16	0,4846
0,29	0,1141	0,74	0,2703	1,19	0,3830	1,64	0,4495	2,18	0,4854
0,30	0,1179	0,75	0,2734	1,20	0,3849	1,65	0,4505	2,20	0,4861
0,31	0,1217	0,76	0,2764	1,21	0,3869	1,66	0,4515	2,22	0,4868
0,32	0,1255	0,77	0,2794	1,22	0,3883	1,67	0,4525	2,24	0,4875
0,33	0,1293	0,78	0,2823	1,23	0,3907	1,68	0,4535	2,26	0,4881
0,34	0,1331	0,79	0,2852	1,24	0,3925	1,69	0,4545	2,28	0,4887
0,35	0,1368	0,80	0,2881	1,25	0,3944	1,70	0,4554	2,30	0,4893
0,36	0,1406	0,81	0,2910	1,26	0,3962	1,71	0,4564	2,32	0,4898
0,37	0,1443	0,82	0,2930	1,27	0,3980	1,72	0,4573	2,34	0,4904
0,38	0,1480	0,83	0,2967	1,28	0,3997	1,73	0,4582	2,36	0,4909

0,39	0,1517	0,84	0,2995	1,29	0,4015	1,74	0,4591	2,38	0,4913
0,40	0,1554	0,85	0,3023	1,30	0,4032	1,75	0,4599	2,40	0,4918
0,41	0,1591	0,86	0,3051	1,31	0,4049	1,76	0,4608	2,42	0,4922
0,42	0,1628	0,87	0,3078	1,32	0,4066	1,77	0,4616	2,44	0,4927
0,43	0,1664	0,88	0,3106	1,33	0,4082	1,78	0,4625	2,46	0,4931
0,44	0,1700	0,89	0,3133	1,34	0,4099	1,79	0,4633	2,48	0,4934
2,50	0,4938	2,64	0,4959	2,78	0,4973	2,92	0,4982	3,60	0,499841
2,52	0,4941	2,66	0,4961	2,80	0,4974	2,94	0,4984	3,80	0,499928
2,54	0,4945	2,68	0,4963	2,82	0,4976	2,96	0,4985	4,00	0,499968
2,56	0,4948	2,70	0,4965	2,84	0,4977	2,98	0,4986	4,50	0,499997
2,58	0,4951	2,72	0,4967	2,86	0,4979	3,00	0,4986		
2,60	0,4953	2,74	0,4969	2,88	0,4980	3,20	0,4993		
2,62	0,4956	2,76	0,4971	2,90	0,4981	3,40	0,4996		

Функция $\Phi(x)$ является нечётной и $\Phi(x) \approx 0{,}5$ при $x \geq 0{,}5$.

Таблица 3. Значения $t_\gamma = t(\gamma;\, n)$

$n \backslash \gamma$	0,95	0,99	0,999	$n \backslash \gamma$	0,95	0,99	0,999
5	2,78	4,60	8,61	19	2,10	2,88	3,92
6	2,57	4,03	6,86	20	2,093	2,861	3,883
7	2,45	3,71	5,96	25	2,064	2,797	3,745
8	2,37	3,50	5,41	30	2,045	2,756	3,659
9	2,31	3,36	5,04	35	2,032	2,729	3,600
10	2,26	3,25	4,78	40	2,023	2,708	3,558
11	2,23	3,17	4,59	45	2,016	2,692	3,527
12	2,20	3,11	4,44	50	2,009	2,679	3,502
13	2,18	3,06	4,32	60	2,001	2,662	3,464
14	2,16	3,01	4,22	70	1,996	2,649	3,439
15	2,15	2,98	4,14	80	1,991	2,640	3,418

16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291

Таблица 4. Критические точки распределения χ^2

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,029
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90

22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Таблица 5. Критические точки распределения Стьюдента

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07

16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,55	3,85
20	1,73	2,09	2,53	2,85	3,55	3,85
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
∞	1,64	1,96	2,33	2,58	3,09	3,29
Число степеней свободы k	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (односторонняя критическая область)					

Таблица 6. Критические точки распределения

F Фишера – Снедекора

(k_1 – число степеней свободы большей дисперсии,

k_2 – число степеней свободы меньшей дисперсии)

Уровень значимости $\alpha = 0{,}01$										
k_1 / k_2	3	4	5	6	7	8	9	10	11	12
3	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11

10	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
Уровень значимости $\alpha = 0{,}05$										
3	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,94
11	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42

Список литературы

1. Айвазян С. А. Теория вероятностей и прикладная статистика / С. А. Айвазян, В. С. Мхитарян. – Москва: ЮНИТИ-ДАНА, 2006. – 656 с.

2. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. – Москва: ФИЗМАТЛИТ, 2006. – 816 с.

3. Кремер Н. Ш. Теория вероятностей и математическая статистика. – Москва: ЮНИТИ, 2002. – 545 с.

4. Гмурман В. Е. Теория вероятностей и математическая статистика. – Москва: Высшая школа, 1977. – 235 с.

5. Xolmurodov E., Yusupo A.I., Aliqulov T.A. Oliy matematika. 3-qism. – Toshkent: “NEXT MEDIA GROUP”, 2017.

6. Xurramov SH. R. Oliy matematika. 2-qism. – Toshkent: “Tafakkur” nashriyoti, 2018.

7. R.N.SHamsiyev. Oliy matematika fanining “Ehtimollar nazariyasi va matematik statistika” qismidan misol va malalar. –T.ToshDTU bosmaxonasi, 2017.

Оглавление