Государственное образовательное учреждение
высшего профессионального образования
СЕВЕРО-ЗАПАДНАЯ АКАДЕМИЯ
ГОСУДАРСТВЕННОЙ СЛУЖБЫ

Курзенев В. А.

ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ДЛЯ УПРАВЛЕНЦЕВ

Учебное пособие

*Рекомендовано Учебно-методическим объединением
по образованию в области статистики и антикризисного
управления в качестве учебного пособия для студентов
высших учебных заведений, обучающихся по специальности
061700 «Статистика», 351000 «Антикризисное управление»
и другим экономическим специальностям*

Издательство СЗАГС
2005

ББК 22.172

Одобрено учебно-методическим советом СЗАГС
Рекомендовано к изданию редакционно-издательским советом СЗАГС

Курзенев В. А. Основы математической статистики для управленцев: Учебное пособие. — СПб: Изд-во СЗАГС, 2005. — 208 с., прил.

Рецензент: Заслуженный работник высшей школы, доктор философских наук, профессор кафедры социологии РГПУ им. А. И. Герцена И. А. Громов.

Учебное пособие содержит основные сведения из теории вероятностей и математической статистики, необходимые для управленцев. Объекты теории вероятностей рассматриваются как математические модели для результатов наблюдений. Основное внимание уделяется случайным величинам и их статистике. Пособие включает элементы теории случайных процессов, многомерного статистического анализа, а также минимально необходимое число примеров и упражнений для усвоения материала и необходимые статистические таблицы для расчетов. Оно адресовано студентам управленческих специальностей блока «Экономика и управление».

ISBN 5-89781-172-5

© СЗАГС, 2005
© Курзенев В. А., 2005: текст

Введение

Большинство задач, встречающихся в практике управления социальными, экономическими и иными системами, связаны с необходимостью учета случайности. Концепция определенного и случайного достаточно подробно обсуждена в философских дисциплинах и здесь не рассматривается. Общеизвестно, что случайное в природе существует так же, как и определенное. В практической деятельности при принятии решений чаще всего приходится иметь дело с результатами наблюдений, выраженных в каких-либо образах или в количественных величинах. Следовательно, результаты наблюдения практически всегда отражают и случайный характер явлений. Если вспомнить слова «отца кибернетики» Н. Винера: «Высшим назначением математики является определение скрытого порядка в хаосе, который нас окружает», то становится понятным, какую роль играет математическая дисциплина, изучающая теорию случая, и сколь важны ее приложения для практики. Эту дисциплину называют теорией вероятностей, и ее основы излагаются в первой части предлагаемой книги. Вторая часть посвящена, по существу, вопросам приложения теории к практике. Ее особенность заключается в том, что в ней излагаются методы использования результатов наблюдений, например, для согласования и проверки их с математическими моделями из теории вероятностей. Классическое название этой части — «Статистика», или, более строго, «Математическая статистика».

Предлагаемое пособие содержит не только основные сведения из теории вероятностей и математической статистики, но и включает, по мнению автора, минимально необходимое число примеров и упражнений для усвоения материала. В то же время оно не претендует на книгу для самостоятельного изучения, хотя и может быть полезным для этих целей. Конспективность изложения делает это пособие не совсем легким для свободного чтения. Однако, как представляется, при внимательной работе с ним можно достаточно оперативно освоить основные понятия и положения теории и получить минимально необходимые навыки для решения практических задач. Книга снабжена основными статистическими таблицами для расчетов. Она адресуется студентам управленческих специальностей блока «Экономика и управление».

Часть I. Вероятность

1. События и вероятность

1.1. Предмет теории вероятностей. Испытания, исходы, события

Математическая наука, изучающая законы случая, называется теорией вероятностей.

Основные объекты изучения: случайные события и случайные величины.

Прежде всего, в теории вероятностей вводятся две взаимосвязанные группы понятий: 1) испытание — опыт — эксперимент — комплекс условий; 2) исходы (результаты), события. Понятия в каждой группе полагают эквивалентными.

Пусть рассматривается некоторый «эксперимент», т. е. осуществляется комплекс условий. Каждый эксперимент — испытание завершается *исходом* или событием.

Примеры испытаний:	*Примеры исходов*:
1) бросание монеты (кости);	1) герб;
2) тасовка карт;	2) раскладка;
3) вынимание карты из колоды;	3) туз;
4) вынимание шара из урны;	4) черный шар;
5) игра в рулетку;	5) 25;
6) выбор наугад;	6) белый шар;
7) «бросание жребия» и т. д.	7) короткий жребий;
8) обнаружение;	8) сигнал;
9) измерение;	9) П=310°;
10) передача информации;	10) сигнал с искажением;
11) страхование.	11) сумма.

Теперь можно дать содержательное определение случайного события.

Определение 1. *Случайным событием A* (событием) по отношению к данному эксперименту S называют всякий факт-исход, который может произойти или не произойти в зависимости от случая.

Для обозначения событий используют большие буквы алфавита: A, B, C ...

Из общего класса событий выделяют события с особыми свойствами.

Определение 2. *Достоверным* событием Ω называют такой исход, который всегда происходит при осуществлении данного эксперимента.

Как будет показано, это событие играет роль единицы при произведении событий.

Определение 3. *Невозможным* событием называют такой исход, который никогда не наступает при осуществлении данного эксперимента.

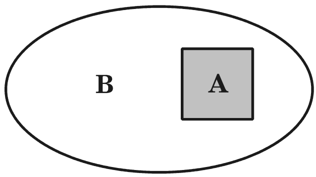
Это событие будет играть роль нуля при сложении событий.

Определение 4. *Противоположным* событием к событию A (или дополнением к A) называют исход-событие, которое происходит только в том случае, если A не происходит.

1.2. Алгебра событий

На случайных событиях необходимо ввести соответствующие операции, чтобы иметь возможность представлять событие через комбинацию более простых. В силу определения события удобной математической моделью для него является множество. Поэтому, как и на множествах, на событиях можно ввести аналогичные операции, т. е. ввести алгебру.

1. Если при появлении A обязательно происходит событие B, то A называют *частным* случаем B ($A \subset B$ или A влечет B).

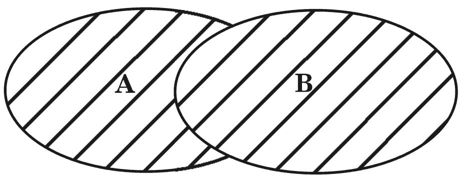


На основании этой операции можно ввести понятие равносильности.

Определение 5. *Равносильными* называются два события, если каждое из них есть частный случай другого, т. е.

$$\left.\begin{matrix} A \subset B \\ B \subset A \end{matrix}\right\} \Rightarrow A = B$$

Суммой двух событий (или объединением) называют событие, равносильное наступлению хотя бы одного из этих двух событий. Используют следующие обозначения: $A \cup B$, $A+B$, A или B.



Для операции суммы (сложения) имеют место следующие соотношения:

1) $A \cup A = A$

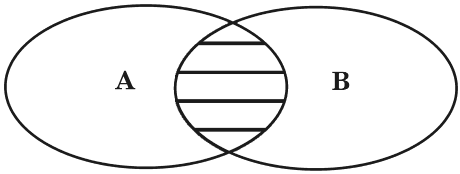
2) $A \cup \Omega = \Omega$

3) $A \cup \varnothing = A$

4) $A \subset B \Rightarrow A \cup B = B$.

Операцию суммирования можно распространить на n событий.

2. *Произведением* (совмещением) двух событий называют событие, равносильное наступлению обоих этих событий. Используются следующие обозначения: $A \cap B$; AB; A и B.



Для операции произведения (умножения) имеют место следующие соотношения:

1) $A \cap A = A$

2) $A \cap \Omega = A$

3) $A \cap \varnothing = \varnothing$

4) $A \subset B \Rightarrow A \cap B = A$.

Операцию произведения можно распространить на n событий.

Для введенных операций суммы и произведения справедливы свойства коммутативности, ассоциативности и дистрибутивности.

$$\left((A \cup B) \cap C = A \cap C \cup B \cap C\right)$$

С помощью произведения вводится понятие несовместных событий.

Определение 6. Два события называются *несовместными*, если их пересечение есть невозможное событие, т. е. $A \cap B = \varnothing$.

3. Разложение события на частные случаи. Если событие представляет собой объединение попарно несовместных событий, т. е. $A = B_1 \cup B_2 \cup \ldots \cup B_k$ и $B_i \bigcap_{i \neq j} B_j = \varnothing$, то говорят, что A распадается на *частные* случаи $B_1, \geq, B_k$.

4. Справедливы следующие соотношения:

1) $A \cup \bar{A} = \Omega$

2) $A \cap \bar{A} = \varnothing$

3) $\overline{A \cup B} = \bar{A} \cap \bar{B}$

4) $A \cup B \cup \bar{A} \cap \bar{B} = \Omega$

5) $(A \cup B) \cap \left(\bar{A} \cap \bar{B}\right) = A \cap \bar{A} \cap \bar{B} \cup \bar{A} \cap B \cap \bar{B} = \varnothing$

6) $\overline{A \cap B} = \bar{A} \cup \bar{B}$

7) $\bar{\Omega} = \varnothing$

8) $\bar{\varnothing} = \Omega.$

В заключение введем два важных определения.

5. **Определение 7.** События B_1, $\geq$, B_k образуют *полную группу событий*, если их объединение есть достоверное событие и они попарно несовместны, т. е.

$B_1 \cup B_2 \cup \ldots \cup B_n = \Omega$ и $B_i \cap B_j = \varnothing$ $i \neq j\, i, j = 1,\, 2,\, \ldots,\, n.$

Пример. $\left\{A, \bar{A}\right\}$ образуют полную группу событий.

Определение 8. Неразложимое событие называют *элементарным*, а полную группу элементарных — *пространством элементарных событий*.

1.3. Определение вероятности

Существует несколько определений вероятности. Наиболее распространенными являются классическое и статистическое.

Определение 9 (классическое). Пусть в результате эксперимента S возможно только n несовместимых и равновозможных исходов $B_1, B_2, \ldots, B_n$, составляющих полную группу событий, и пусть A есть сумма определенных m из них, т. е. $A = B_{i_1} \cup B_{i_2} \cup \ldots. \cup B_{i_m}$, тогда *вероятностью* события A называют число, равное отношению числа m к числу n, т. е. $P(A) \equiv \dfrac{m}{n}$.

Замечание 1. Это определение применимо к комбинаторным задачам, когда число исходов конечно и события равновозможны.

Определение 10 (статистическое). Пусть S — эксперимент из n *независимых*, осуществляемых на практике испытаний и пусть событие A наступает μ раз, тогда число $P(A)$, около которого колеблется *частота* события называют *статистической вероятностью,* т. е. $P(A)=\lim\limits_{n\to\infty}\frac{\mu}{n}$.

Замечание 2. Число n может быть очень большим, и найти предел не всегда оказывается возможным.

Замечание 3. Статистическое определение не использует предположения о равновозможности исходов, в чем состоит его достоинство, однако оно использует схему независимых испытаний на практике, что связано с определенными затратами, и с учетом замечания 2 также имеет ограничения.

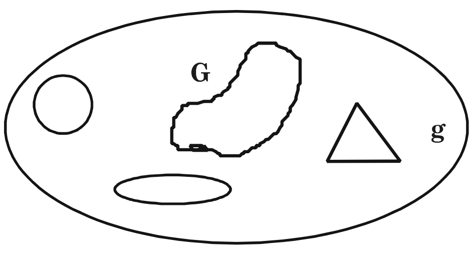
Существуют и другие определения. Приведем еще одно.

Определение 11 (геометрическое). Пусть на множество из n-мерного пространства $G\subset R^n$ бросается n-мерная точка наугад и пусть имеется подмножество $g\in G$, тогда под вероятностью попадания этой точки на указанное подмножество понимают отношение меры подмножества к мере всего множества, т. е.

$$P\big(\text{попад }(\cdot)\text{ в }g\big)=\frac{mesg}{mesG}.$$

Замечание 4. В определении 11 имеет место бесконечное множество равновозможных исходов. При этом здесь равновозможные события — события, когда точки попадают в области равной меры, например, площади.

Пример. Задача Бюффона, игра в рулетку.



1.4. Примеры и задачи

Примеры

Пример 1.1. Имеются две игральные кости. Событие A состоит в том, что сумма выпавших чисел четна. Событие B — на одной из граней единица. Каков смысл событий: 1) $A \bigcup B$; 2) $A \bigcap B$; 3) $\overline{B}$; 4) $A \bigcap B$?

Решение:

1) Сумма выпавших чисел четна или же нечетна, но на одной из костей единица, а на другой четное число.
2) На одной из костей единица, а на другой нечетное число.
3) Ни на одной из костей нет единицы.
4) Четные суммы кроме пар с единицей.

Пример 1.2. Событие A — хотя бы один из трех проектов не окупается. Событие B — все три проекта окупаются. Что означают события: 1) $A \bigcap B$; 2) $A \bigcup B$?

Решение:

1) События A и B не могут происходить одновременно, поэтому $A \bigcap B$ есть невозможное событие $\varnothing$.
2) Событие A или B всегда имеет место, поэтому $A \bigcup B$ есть достоверное событие.

Пример 1.3. В группе 25 студентов, из них 10 юношей и 15 девушек. Какова вероятность, что наугад вызванный студент будет юношей?

Решение:

A — событие, что наугад вызванный студент — юноша. Вероятность $P(A) = m/n$, где $m = 10$, а $n = 25$, тогда $P(A) = 0{,}4$.

Пример 1.4. Из 30 карточек с различными буквами русского алфавита наугад одну за другой берут 7 карточек. Какова вероятность, что эти буквы в порядке их вынимания составят слово «СПУТНИК»?

Решение:

Число благоприятных исходов $m = 1$, число всех равновозможных исходов:

$$n = C_{30}^{7} = 30!/7!(30-7)! = 24 \cdot 25 \cdot 26 \cdot 27 \cdot 28 \cdot 29 \cdot 30 \approx 10^{10}$$

$$P(A) = 1/10^{10} = 10^{-10}.$$

Пример 1.5. Из колоды карт (52) вынимаются наугад сразу 3 карты. Какова вероятность, что это будут «тройка», «семерка», «туз»?

Решение:

Число троек, семерок и тузов — по четыре (по числу мастей), т. е. благоприятных исходов $m = 4 \cdot 4 \cdot 4 = 64$, а число всех различных троек карт как равновозможных исходов тогда

$P(A) = 64 \cdot 3! \cdot 49!/52! = 0{,}0029.$

Задачи

1.1. Пусть имеется множество брачных пар. Событие A состоит в том, что «жениху больше 20 лет», событие B: «жених старше невесты», событие C: «невесте больше 20 лет». Каков смысл событий: 1) $A \cap B \cap C$; 2) $\bar{B}$; 3) $A \cap \bar{B} \cap C$?; 4) $A \cap \bar{C}$?

Ответ:

а) «жених старше невесты и обоим больше 20 лет»,

б) «жених не старше невесты»;

в) «жениху и невесте больше 20 лет и жених не старше невесты»;

г) «жениху больше, а невесте не больше 20 лет».

1.2. Пусть A, B, C — произвольные события. Упростить следующие выражения для событий:

а) $(A \cup B) \cap (A \cup \bar{B})$;

б) $(A \cup B) \cap (B \cup C)$;

в) $(A \cup B) \cap (A \cup \bar{B})(\bar{A} \cup B)$.

Ответ: а) A; б) $B \cup A \cap C$; в) $(A \cap B)$.

1.3. A, B, C – произвольные события. Как записать, что произошли: а) все события, б) ни одного события, в) только событие A, г) только A и B, д) по крайней мере одно, е) по крайней мере два, ж) одно и только одно событие?

Ответ:

а) $A \cap B \cap C$; б) $\bar{A} \cap \bar{B} \cap \bar{C}$; в) $\mathcal{A} \cap \bar{\mathcal{B}} \cap \bar{\mathcal{C}}$; г) $\mathcal{A} \cap \mathcal{B} \cap \bar{\mathcal{C}}$;

д) $A \cup B \cup C$; е) $(A \cap B) \cup (A \cap C) \cup (B \cap C)$;

ж) $(\mathcal{A} \cap \bar{\mathcal{B}} \cap \bar{\mathcal{C}}) \cup (\mathcal{B} \cap \bar{\mathcal{A}} \cap \bar{\mathcal{C}}) \cup (\mathcal{C} \cap \bar{\mathcal{B}} \cap \bar{\mathcal{A}})$.

1.4. Тома пятитомной энциклопедии по экономике стоят на полке в случайном порядке. Какова вероятность, что тома окажутся упорядоченными?

Ответ: $P(A)=2/5!=1/60$.

1.5. Телефонный номер состоит из пяти цифр. Какова вероятность, что цифры номера разные?

Ответ: $P(A)=C_{10}^{5}/10^{5}=0{,}3024$.

1.6. Группа студентов в 28 человек с соотношением юношей и девушек 1:1 на практических занятиях делится пополам. Найти вероятность, что в подгруппах число юношей и девушек одинаково.

Ответ: $P(A)=\left(C_{14}^{7}\cdot C_{14}^{7}\right)/C_{28}^{14}\approx 0{,}59$.

1.7. Среди 16 экзаменационных билетов 4 содержат относительно легкие вопросы. Определить вероятность: а) первому экзаменующемуся взять один из таких билетов; б) двум экзаменующимся не взять ни одного такого билета.

Ответ: а) $P(A)=\dfrac{4}{16}=0{,}25$;

б) $P(B)=\dfrac{C_{12}^{2}}{C_{16}^{2}}=\dfrac{12!\cdot 2!\cdot 14!}{2!\cdot 10!\cdot 16!}=\dfrac{11}{20}\approx 0{,}55$

1.8. Используя условия задачи 1.7, определить вероятность того, что из четырех первых экзаменующихся двое возьмут билеты с относительно легкими вопросами.

Ответ: $\mathcal{P}(\mathcal{A})=\dfrac{C_{4}^{2}\cdot C_{12}^{2}}{C_{16}^{4}}=\dfrac{99}{455}\approx 0{,}217$.

1.9. Полная колода карт (52 карты) делится пополам. Найти вероятность того, что число черных и красных карт в обеих пачках будет одинаковым.

Ответ: $\mathcal{P}(\mathcal{A}) = \dfrac{\left(C_{26}^{13}\right)^2}{C_{52}^{26}} = \dfrac{(26!)^4}{(13!)^4 \cdot 52!} \approx 0{,}22.$

1.10. Среди 25 студентов группы, в которой 10 девушек, разыгрывается 5 билетов. Определить вероятность того, что среди обладателей билетов окажутся две девушки.

Ответ: $\mathcal{P}(\mathcal{A}) = \dfrac{C_{15}^3 \cdot C_{10}^2}{C_{25}^5} = \dfrac{195}{506} \approx 0{,}385.$

1.11. В фирму поступило 15 компьютеров. Известно, что шесть из них нуждаются в общей регулировке. Мастер берет первые попавшиеся пять. Какова вероятность того, что два из них нуждаются в общей регулировке?

Ответ: $\mathcal{P}(\mathcal{A}) = \dfrac{C_9^3 \cdot C_6^2}{C_{15}^5} = \dfrac{60}{143}.$

1.12. В лифт девятиэтажного дома на первом этаже вошли пять человек. Каждый из них с одинаковой вероятностью выходит на любом этаже, начиная со второго. Найти вероятность того, что: а) все пассажиры выйдут на одном и том же этаже; б) все выйдут на шестом этаже; в) все выйдут на разных этажах.

Ответ: а) $\mathcal{P}(\mathcal{A}) = \dfrac{8}{8^5} = \dfrac{1}{8^4} = \dfrac{1}{4096}$; б) $\mathcal{P}(\mathcal{B}) = \dfrac{1}{8^5} = \dfrac{1}{32768}$;

в) $\mathcal{P}(\mathcal{C}) = \dfrac{56}{8^5} = \dfrac{7}{4096}.$

1.13. В условии задачи 1.12 найти вероятность того, что на одном из этажей выйдут три человека, а на другом — два.

Ответ: $\mathcal{P}(\mathcal{A}) = \dfrac{C_5^3 \cdot C_8^1 \cdot C_7^1}{8^5}.$

2. Свойства вероятности. Аксиомы теории вероятностей

2.1. Основные свойства вероятности

Нетрудно заметить, что приведенные в п. 1.3 определения вероятности связаны с определенной схемой и несут в некотором смысле печать конкретности. Естественно стремление иметь единое определение вероятности. Это можно сделать, если перейти на более высокий уровень абстракции, для чего необходимо сначала рассмотреть свойства вероятностей, определенных различным образом.

А. Свойства вероятности согласно определению 9 (классическому):

1) Вероятность есть число, заключенное между нулем и единицей, т. е. $0 \le P(A) \le 1$.

   В самом деле это так. Если событие A есть невозможное, то имеем $A = \varnothing \Rightarrow P(A) = \dfrac{0}{n} = 0$, а если событие A есть достоверное, то имеем $A = \Omega = B_1 \cup B_2 \cup \ldots \cup B_n \Rightarrow P(A) = P(\Omega) = \dfrac{n}{n} = 1$.

   Для других событий вероятность будет заключаться между этими двумя крайними значениями.

2) Вероятность достоверного события всегда равна единице, $P(\Omega) = 1$.

   Это показано выше.

3) Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий, т. е. $A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$.

   В самом деле, из n несовместимых равновозможных исходов событию A соответствует m_1 из них, а событию B — m_2 из них. Тогда объединению $A \cup B$ соответствует $m_1 + m_2$ исходов из n. По определению 9 имеем

   $P(A \cup B) = \dfrac{m_1 + m_2}{n} = \dfrac{m_1}{n} + \dfrac{m_2}{n} = P(A) + P(B)$ — *формула сложения вероятности*.

4) Вероятность противоположного события равна 1 минус вероятность исходного, т. е. $P(\bar{A}) = 1 - P(A)$.

   Это действительно так, поскольку

   $$P(\bar{A}) = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(A).$$

   Или $A \cup \bar{A} = \Omega$; $P(A \cup \vec{A}) = P(\Omega)$; $P(A) + P(\bar{A}) = 1$.

5) Вероятность частного случая не превосходит вероятность всего события, т. е. $A \subset B \Rightarrow P(B) \geq P(A)$.

В самом деле, $B = A \bigcup \bar{A} \bigcap B$. $P(B) = P(A) + P(\bar{A} \bigcap B) \geq P(A)$.

В. Свойства вероятности согласно определению 10 (статистическому): свойства этой вероятности очевидно определяются свойствами частоты.

Частота события $w(A) = \dfrac{\mu}{n}$ характеризуется следующим:

1) Частота события есть число, заключенное между нулем и единицей, т. е. $0 \leq w(A) \leq 1$, поскольку для невозможного события m равно нулю, а для достоверного — n.

2) Частота достоверного события равна единице, т. е. $w(\Omega) = 1$, т. к. для него всегда $\mu = n$.

3) Частота суммы двух несовместных событий равна сумме частот этих событий, т. е.

   $$A \bigcap B = \varnothing \Rightarrow w(A \bigcup B) = w(A) + w(B),$$

   т. к. $w(A \bigcup B) = \dfrac{\mu_A + \mu_B}{n}$.

4) Частота противоположного события равна разности между единицей и исходным событием. Это можно записать как

   $$w(\bar{A}) = 1 - w(A),$$

   т. к. $A \leftrightarrow \mu_A$, $\bar{A} \leftrightarrow n - \mu_A \Rightarrow w(\bar{A}) = \dfrac{n - \mu_A}{n} = 1 - w(A)$.

Поскольку по определению 10 $P(A)=\lim\limits_{n\to\infty} w(A)$, то очевидно, что свойства статистической вероятности $P(A)$ аналогичны свойствам частоты (можно этого даже потребовать). Следовательно, для статистической вероятности имеют место те же свойства:

1) $0 \le P(A) \le 1$;
2) $P(\Omega)=1$;
3) $A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$ — *формула сложения вероятностей*;
4) $P(\bar{A}) = 1 - P(A)$.

Замечание 1. Очевидно формальное совпадение свойств 1–4 вероятности определений 9 и 10. В математике показывается и их совпадение со свойствами вероятности определения 11. Все это дает основание перейти от рассмотренных частных определений к общему абстрактному определению вероятности как некоторой функции, заданной на множестве.

2.2. Вероятностное пространство

Построим некоторую формальную конструкцию. Из определения 8 п. 1.2 пространство элементарных событий есть полная группа неразложимых событий, сумма которых есть достоверное событие $\Omega = \{B_i\}$; $i = 1, 2, \ldots, n$ при их попарной несовместности, т. е. $B_i \cap B_j = \varnothing$, $i \ne j$. Далее, пусть рассматривается некоторая *система подмножеств F* этих элементарных событий B_i. Элементы этой системы F будем называть случайными событиями. Другими словами, случайное событие есть некоторая комбинация элементарных событий (объединение, пересечение). Пусть система подмножеств такова, что она включает сумму, произведение событий, противоположное, достоверное и невозможное событие, т. е.

1) $A, B \in F \Rightarrow A \cup B \in F$;
2) $A, B \in F \Rightarrow A \cap B \in F$;

3) $A \in F \Rightarrow \bar{A} \in F$;

4) $\Omega \in F$;

5) $\varnothing \in F$.

Следовательно, эта система представляет собой множество всевозможных событий, порожденных пространством элементарных, и содержит достоверное и невозможное события. Такая система имеет специальное название.

Определение 1. Система подмножеств F со свойствами 1—5 называется σ-алгеброй или борелевским телом подмножеств.

Теперь можно перейти к следующему определению.

Определение 2. Пусть для *любого* события $A \in F$ определено неотрицательное число, т. е. задана числовая функция $P(A)$ со свойствами вероятности, тогда тройка $\{\Omega, F, P\}$ называется *вероятностным пространством*, а числовая функция множества P, определенная на σ-алгебре F, *вероятностью*.

2.3. Аксиоматика теории вероятностей

Введение функции P на F должно сопровождаться введением некоторых правил — аксиом с учетом свойств вероятности.

Аксиома 1. Пусть имеется пространство элементарных событий $\Omega = \{B_i\}$ и пусть F — σ-алгебра, тогда любой элемент A из σ-алгебры F есть *случайное событие*.

Аксиома 2. Для любого события A из σ-алгебры F существует числовая функция P со значениями между нулем и единицей, т. е. $\forall A \in F\ \exists P(A):\ 0 \le P(A) \le 1$, которую называют *вероятностью*.

Аксиома 3. Значение функции от достоверного события равно единице, т. е. $P(\Omega) = 1$.

Аксиома 4. Значение функции для двух непересекающихся событий равно сумме значений функций от этих двух событий, т. е.

$$A, B \in F;\ A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B).$$

Значение функции от суммы n непересекающихся событий равно сумме значений этой функции от событий, т. е.

$$\{A_n\} \in F;\ A_i \cap A_j = \varnothing \Rightarrow P\left(\bigcup_n A_n\right) = \sum_n P(A_n).$$

Эту аксиому называют аксиомой конечной, а если $n \to \infty$ — счетной аддитивности.

Аксиома 5. Если имеется система вложенных событий с числом, стремящимся к бесконечности, и такая, что произведение их стремится к невозможному событию, то предел последовательности значений функций равен нулю, т. е.

$$A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots \text{ и } \bigcap_{n=1}^{\infty} A_n = \varnothing \Rightarrow \lim_{n\to\infty} P(A_n) = 0.$$

Эту аксиому называют аксиомой непрерывности.

При введении аксиом необходимо, чтобы система аксиом была непротиворечива. Указанная система удовлетворяет этому требованию.

Замечание 2. Система F является полной согласно аксиоме 5, а система аксиом 1–6 имеет некоторый произвол в задании функции P, пространстве Ω и в этом смысле не является полной.

Примеры вероятностных пространств:

Пример 1. Игра с бросанием монеты.

$$\Omega : \{\mathcal{J}.\mathcal{P}.\}$$

$$F : \{\varnothing, \mathcal{J}, \mathcal{P}, \mathcal{J} \cup \mathcal{P} = \Omega\}$$

$$P : P(\mathcal{J}) = \mathcal{P}(\mathcal{P}) = \frac{1}{2}.$$

Пример 2. Бросание кости.

$$\Omega : \{1.2.3.4.5.6\}$$

$$F : \{\text{множество подмножеств } \Omega\}$$

$$P : \{P_1.P_2.P_3.P_4.P_5.P_6\}$$

$$P_i \geq 0 \text{ и } P_1 + \ldots + P_6 = 1$$

$$A = \{i_1, i_2, \ldots, i_k\}_{k\leq 6} \to P(A) = P_{i_1} + P_{i_2} + \ldots + P_{i_k},\ k \leq 6.$$

3. Основные теоремы теории вероятностей

3.1. Условная вероятность

На возможность осуществления некоторого события A может влиять наступление некоторого другого события B.

Эта связь выражается через так называемую *условную вероятность*.

Определение 1. Пусть A и B — случайные события и пусть вероятность наступления события B строго больше нуля, т. е. $P(B)>0$. Тогда вероятность наступления события A при условии осуществления события B называют *условной* вероятностью, обозначают через $P(A|B)$ или $P_B(A)$ и определяют по правилу:

$$P_B(A)=\frac{P(A\cap B)}{P(B)}.$$

Другими словами, *условной вероятностью* события A при условии события B называют отношение $P(A\cap B)$ к $P(B)$.

Пример.

$$\left.\begin{array}{l} A\leftrightarrow m \text{ благоприятных исходов} \\ B\leftrightarrow l \text{ благоприятных исходов} \\ A\cap B\leftrightarrow k \text{ благоприятных исходов} \end{array}\right\}\leftrightarrow P_B(A)=\frac{k}{l}=$$

$$=\frac{k}{n}\Big/\frac{l}{n}=\frac{P(A\cap B)}{P(B)}.$$

Введенная функция удовлетворяет основным аксиомам теории вероятностей и поэтому принимается за вероятность. Это легко проверить:

1) $P_B(A)$ — есть функция от A и для любого A неотрицательна, $P_B(A)\geq 0$;

2) $0\leq P_B(A)\leq 1$, поскольку $A\cap B\subset B\Rightarrow P(A\cap B)\leq P(B)$;

3) $P_B(\Omega)=1,\ P_B(\Omega)=\dfrac{P(\Omega\cap B)}{P(B)}=\dfrac{P(B)}{P(B)}=1,$

что следует из $\Omega\cap B=B$;

4) $A\cap C=\varnothing$, тогда $P_B(A\cup C)=\dfrac{P\big((A\cup C)\cap B\big)}{P(B)}=$

$$=\frac{P(A\cap B)+P(C\cap B)}{P(B)}=P_B(A)+P_B(C),$$

$$P_B\left(\bigcup_n A_n\right)=\frac{P\left(\bigcup_n (A_n\cap B)\right)}{P(B)}=\sum_n\frac{P(A_n\cap B)}{P(B)}=\sum_n P_B(A_n).$$

Замечание 1. Можно рассматривать другую условную вероятность $P_A(B)=\dfrac{P(A\cap B)}{P(A)}$. Сопоставление $P_A(B)$ с $P_B(A)$ приводит к формуле умножения вероятностей:

$$\boxed{P(A\cap B)=P(B)\cdot P_B(A)=P(A)\cdot P_A(B)}$$

Пример. В урне находятся a белых шаров и b черных шаров. Какова P (вероятность вытаскивания двух черных шаров подряд)?

B — первое вытаскивание черного шара;

A — второе вытаскивание черного шара.

$$P(A\cap B)=P(B)\cdot P_B(A)=\frac{b}{a+b}\cdot\frac{b-1}{a+b-1}.$$

3.2. Формулы полной вероятности и Байеса

Теорема 1 (формула полной вероятности).

Вероятность произвольного события, зависящего от событий, составляющих полную группу, равна сумме произведений вероятностей этих событий на условные вероятности исходного события. Это можно записать так:

$$\left.\begin{array}{l}\text{Если: 1) } A_1,\ A_2,\ldots,\ A_n \text{ — полная группа событий} \\ \qquad A_i \cap A_j = \varnothing \text{ при } i \neq j;\ A_1 \cup A_2 \cup \ldots \cup A_n = \Omega \\ \quad \text{2) } B \text{ — произв. сл. (завис. от } A_i\text{)}\end{array}\right\} \Rightarrow P(B) =$$

$$= \sum_{i=1}^{n} P(A_i) P_{A_i}(B) \text{ — формула полной зависимости.}$$

Доказательство.

$\{A_i\}$ — полная группа, поэтому имеем

$$\bigcup_{i=1}^{n} A_i = \Omega;\ B \cap \Omega = B \cap \left(\bigcup_{i=1}^{n} A_i \right) = \bigcup_{i=1}^{n} A_i \cap B.$$

Поскольку $A_i \cap A_j = \varnothing$ при $i \neq j$, то и $(A_i \cap B) \cap (A_j \cap B) = \varnothing$.

Тогда с учетом аксиомы аддитивности имеем

$$P(B \cap \Omega) = P(B) = P \bigcup_{i=1}^{n} (A_i \cap B) = \sum_{i=1}^{n} P(A_i \cap B).$$

Но согласно теореме умножения $P(A_i \cap B) = P(A_i) \cdot P_{A_i}(B)$.

Следовательно, $P(B) = \sum\limits_{i=1}^{n} P(A_i) P_{A_i}(B)$.

Замечание 2. A_i называют иногда *гипотезами*.

Теорема 2 (формула Байеса).

При условиях теоремы 1 справедлива формула для определения вероятности выполнения событий, составляющих полную группу, т. е. гипотез, при наступлении произвольного события:

$$\boxed{P_B(A_i) = \frac{P_{A_i}(B) \cdot P(A_i)}{\sum_{i=1}^{n} P(A_i) P_{A_i}(B)}}$$

Доказательство. По определению 1 и формуле умножения вероятностей имеем

$$P_B(A_i) = \frac{P(A_i \cap B)}{P(B)} = \frac{P_{A_i}(B) \cdot P(A_i)}{P(B)} \overset{\text{т. 1}}{=} \frac{P_{A_i}(B) \cdot P(A_i)}{\sum_{i=1}^{n} P(A_i) P_{A_i}(B)}.$$

Замечание 3. Формула Байеса дает ответ на вопрос: какова вероятность, что при наступлении события B наступит A_i.

3.3. Независимость случайных событий. Теорема сложения

Определение 2. События A и B называются *независимыми*, если вероятность наступления каждого из них не зависит от наступления другого, т. е.

$$P_B(A) = P(A), \text{ откуда } P(A \cap B) = P(A) \cdot P(B).$$

Теорема 3. Если два события независимы, то они независимы и с противоположными и независимы их противоположные. Это можно записать иначе:

$$A, B \text{ — независимы } \Rightarrow \left.\begin{array}{r} 1)\ A \text{ и } \bar{B} \\ 2)\ \bar{A} \text{ и } B \\ 3)\ \bar{A} \text{ и } \bar{B} \end{array}\right\} \text{ независимы.}$$

Доказательство (для 1).

$P(A\cap B)=P(A)P(B)$ — по определению 2 и условию теоремы,

$P\left(A\cap \bar{B}\right)=P(A)P_A\left(\bar{B}\right)$ — по теореме умножения, но

$$P(A)=P(A)P(B)+P(A)P_A\left(\bar{B}\right)\therefore 1=P(B)+P_A\left(\bar{B}\right)\Rightarrow P\left(\bar{B}\right)=P_A\left(\bar{B}\right),$$

что и требовалось доказать.

Замечание 4. Так как теорема умножения может быть распространена на n событий

$$P(A_1,A_2,\ldots,A_n)=P(A_1)\cdot P_{A_1}(A_2)\cdot P_{A_1A_2}(A_3)\cdots P_{A_1\cdots A_{n-1}}(A_n),$$

то для независимых событий $P(A_1,A_2,\ldots,A_n)=\prod_{i=1}^{n}P(A_i)$.

Теорема 4. Вероятность суммы двух произвольных событий равна сумме вероятностей этих событий за вычетом вероятности произведения событий, т. е.

$$P(A\cup B)=P(A)+P(B)-P(A\cap B).$$

Доказательство. В самом деле, имеем

$$A=(A\cap B)\cup\left(A\cap\bar{B}\right)\leftrightarrow P(A)=P(A\cap B)+P\left(A\cap\bar{B}\right)$$

$$B=(A\cap B)\cup\left(\bar{A}\cap B\right)\leftrightarrow P(B)=P(A\cap B)+P\left(\bar{A}\cap B\right)$$

$$A\cup B=\left(A\cap\bar{B}\right)\cup\left(\bar{A}\cap B\right)\cup(A\cap B)\leftrightarrow P(A\cup B)=$$

$$=P\left(A\cap\bar{B}\right)+P\left(\bar{A}\cap B\right)+P(A\cap B)$$

$$P(A)+P(B)=P(A\cap B)+\underbrace{P\left(A\cap\bar{B}\right)+P\left(\bar{A}\cap B\right)+P(A\cap B)}_{\substack{/\!/\\ P(\mathcal{A}\cup\mathcal{B})}}\therefore$$

$$P(A\cup B)=P(A)+P(B)-P(A\cap B).$$

Можно показать, что для трех произвольных событий:

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-$$
$$-P(B\cap C)+P(A\cap B\cap C).$$

Для независимых событий: $P\left(\bigcup\limits_{i=1}^{n} A_i\right)=1-\prod\limits_{i=1}^{n} P\left(\bar{A}_i\right).$

3.4. Примеры и задачи

Примеры

Пример 3.1. Бросаются две игральные кости. Какова вероятность того, что произведение выпавших чисел будет: а) четным; б) кратным трем?

Решение:

а) A — событие на 1-й кости — четное число; событие B — на 2-й кости – четное число, $A \cup B$ — хотя бы на одной четное число. Тогда $P(A \cup B)=P(A)+P(B)-P(A \cap B)$;

$$\mathcal{P}(\mathcal{A})=\mathcal{P}(\mathcal{B})=\frac{3}{6}=\frac{1}{2} \Rightarrow \mathcal{P}(\mathcal{A} \cup \mathcal{B})=\frac{1}{2}+\frac{1}{2}-\frac{1}{2}\cdot\frac{1}{2}=\frac{3}{4}=0{,}75.$$

б) $\mathcal{P}(\mathcal{A})=\mathcal{P}(\mathcal{B})=\frac{2}{6}=\frac{1}{3} \Rightarrow \mathcal{P}(\mathcal{A} \cup \mathcal{B})=\frac{1}{3}+\frac{1}{3}-\frac{1}{3}\cdot\frac{1}{3}=\frac{5}{9}.$

Пример 3.2. На экспертизу поступил проект решения, который был направлен одновременно трем экспертам. В проекте имеется противоречие. Вероятность его выявления первым экспертом 0,2; вторым — 0,5; третьим — 0,6. Какова вероятность, что оно будет выявлено: а) только первым экспертом; б) хотя бы одним экспертом; в) вторым и третьим экспертами?

Решение:

а) Пусть A_1 — событие обнаружил первый эксперт, A_2 — второй и A_3 – третий. Тогда $\mathcal{A}_1 \cap \bar{\mathcal{A}}_2 \cap \bar{\mathcal{A}}_3$ есть оцениваемое событие A и $\mathcal{P}(\mathcal{A})=\mathcal{P}\left(\mathcal{A}_1 \cap \bar{\mathcal{A}}_2 \cap \bar{\mathcal{A}}_3\right)=0{,}2\cdot 0{,}5\cdot(1-0{,}6)=0{,}04.$

б) Пусть $A_1 \cup A_2 \cup A_3 = A$ — оцениваемое событие.

Но $\left(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3\right) \cup \bar{\mathcal{A}}=\Omega$, $\bar{\mathcal{A}}=\overline{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3}=\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2 \cap \bar{\mathcal{A}}_3$

$$P\left(A\cup\bar{A}\right)=P\left(A\right)+P\left(\bar{A}\right)=1\Rightarrow P\left(A\right)=1-P\left(\bar{A}_1\right)\cdot P\left(\bar{A}_2\right)\cdot P\left(\bar{A}_3\right)=$$
$$=1-(1-0{,}2)\cdot(1-0{,}5)\cdot(1-0{,}6)=0{,}84.$$

в) $\bar{\mathcal{A}}_1\cap\mathcal{A}_2\cap\mathcal{A}_3=\mathcal{A}$ — оцениваемое событие.

$$\mathcal{P}\left(\mathcal{A}\right)=\left(1-\mathcal{P}\left(\mathcal{A}_1\right)\right)\cdot\left(\mathcal{P}\mathcal{A}_2\right)\cdot\mathcal{P}\left(\mathcal{A}_3\right)=(1-0{,}2)\cdot0{,}5\cdot0{,}6=0{,}24.$$

Пример 3.3. Наземная система общественного транспорта города состоит из трех видов транспорта: автобусов, троллейбусов и трамваев. Автобусы составляют 50% всего парка, троллейбусы — 30%, трамваи — 20%. Доля эксплуатируемых машин, не соответствующих требованиям по техническому состоянию, составляет соответственно 0,025; 0,020 и 0,015. Какова вероятность, что взятое наугад транспортное средство полностью исправно?

Решение:

Пусть A_i – взятое наугад транспортное средство и относится к i-му виду (i=1, 2, 3), а B — транспортное средство исправно. Тогда $P(A_1)$ = 0,5; $P(A_2)$ = 0,3; $P(A_3)$ = 0,2. Условные вероятности

$$\mathcal{P}_{\mathcal{A}_1}\left(\mathcal{B}\right)=1-0{,}025=0{,}975,\quad \mathcal{P}_{\mathcal{A}_2}\left(\mathcal{B}\right)=1-0{,}020=0{,}980,$$

$$\mathcal{P}_{\mathcal{A}_3}\left(\mathcal{B}\right)=1-0{,}015=0{,}985,$$

$$\mathcal{P}\left(\mathcal{B}\right)=\mathcal{P}\left(\mathcal{A}_1\right)\cdot\mathcal{P}_{\mathcal{A}_1}\left(\mathcal{B}\right)+\mathcal{P}\left(\mathcal{A}_2\right)\cdot\mathcal{P}_{\mathcal{A}_2}\left(\mathcal{B}\right)+\mathcal{P}\left(\mathcal{A}_3\right)\cdot\mathcal{P}_{\mathcal{A}_3}\left(\mathcal{B}\right)=$$
$$=0{,}5\cdot0{,}975+0{,}3\cdot0{,}980+0{,}2\cdot0{,}985\approx0{,}98.$$

Пример 3.4. Точки и тире при телеграфе встречаются в отношении 5:3. При передаче $2/5$ числа «точек» и $1/3$ числа «тире» искажаются. Какова вероятность, что на выходе будет точка?

Решение:

A_1 — передана «точка», A_2 — передано «тире», B — на выходе «точка».

$$\mathcal{P}\left(\mathcal{A}_1\right)=\frac{5}{8};\ \mathcal{P}\left(\mathcal{A}_2\right)=\frac{3}{8};\mathcal{P}_{\mathcal{A}_1}\left(\mathcal{B}\right)=1-\frac{2}{5}=\frac{3}{5};\ \mathcal{P}_{\mathcal{A}_2}\left(\mathcal{B}\right)=\frac{1}{3};$$
$$\mathcal{P}\left(\mathcal{B}\right)=\frac{5}{8}\cdot\frac{3}{5}+\frac{3}{8}\cdot\frac{1}{3}=\frac{1}{2}.$$

Пример 3.5. Два юриста рассматривают документ независимо друг от друга, делая по одной правке. Квалификация первого юриста 0,8, а второго — 0,4. Сделана одна верная правка. Найти вероятность, что верную правку сделал первый юрист.

Решение:

A_1 — оба юриста сделали по неверной правке; A_2 — оба юриста сделали по верной правке; A_3 — у первого юриста верная правка, у второго — неверная; A_4 — у первого неверная правка, у второго верная.

$$P(A_1)=(1-0{,}8)(1-0{,}4)=0{,}12;\ P(A_2)=0{,}8\cdot 0{,}4=0{,}32;$$

$$P(A_3)=0{,}8(1-0{,}4)=0{,}48;\ P(A_4)=(1-0{,}8)\cdot 0{,}4=0{,}08.$$

B – одна верная правка.

$$P_{н_1}(B)=0;\ P_{н_2}(B)=0;\ P_{н_3}(B)=1;\ P_{н_4}(B)=1;$$

$$P_e(A_1)=\frac{0{,}48\cdot 1}{0{,}12\cdot 0+0{,}32\cdot 0+0{,}48\cdot 1+0{,}08\cdot 1}=\frac{6}{7}.$$

Задачи

3.1. Курсант сдаст зачет по стрельбе, если получит оценку не ниже 4. Какова вероятность сдачи зачета, если известно, что курсант получает за стрельбу 5 с вероятностью 0,3 и 4 с вероятностью 0,6?

Ответ: 0,9.

3.2. В коробке лежат 8 красных и 12 синих карандашей. Наудачу вынимают три карандаша. Какова вероятность того, что хотя бы один из них красный?

Ответ: $\approx 0{,}8$.

3.3. Документ последовательно печатается четырьмя машинистками независимо друг от друга. Вероятность внесения ошибки каждой машинисткой равна 0,01. Какова вероятность выпуска документа без ошибок?

Ответ: $\approx 0{,}96$.

3.4. При включении двигатель начинает работать с вероятностью 0,8. Какова вероятность того, что он начнет работать со второго включения?

Ответ: 0,16.

3.5. Монету бросают до тех пор, пока не появится два герба или две решки. Найти вероятность того, что потребуется не более трех бросаний.

Ответ: 0,75.

3.6. Четыре охотника стреляют одновременно и независимо друг от друга по буйволу. Какова вероятность того, что буйвол подстрелен, если вероятность попадания для каждого охотника равна 2/3?

Ответ: $\frac{80}{81}$.

3.7. Часы изготавливаются на трех заводах и поступают в магазин. Первый завод производит 40% поступающей продукции, второй — 45%, третий — 15%. В продукции первого завода спешат 80% часов, у второго — 70%, у третьего — 90%. Какова вероятность того, что купленные в этом магазине часы спешат?

Ответ: 0,77.

3.8. Группа студентов состоит из пяти отличников, шести хорошо успевающих и двенадцати занимающихся слабо. Отличники на экзамене могут получить только отличные оценки. Хорошо успевающие студенты могут получить с равной вероятностью хорошие и отличные оценки. Слабо занимающиеся могут получить с равной вероятностью хорошие, удовлетворительные и неудовлетворительные оценки. Для сдачи экзамена наугад вызывается один студент. Найдите вероятность того, что он получит хорошую или отличную оценку.

Ответ: $\frac{15}{23}$.

3.9. Из десяти монет четыре поддельные. Поддельная монета легче нормы с вероятностью 0,3, а неподдельная монета легче нормы с вероятностью 0,1. Взятая наудачу монета оказалась легче нормы. Найдите вероятность того, что она поддельная.

Ответ: $\frac{2}{3}$.

3.10. Три стрелка одновременно выстрелили, и в мишени обнаружены две пули. Найти вероятность того, что третий стрелок поразил мишень, если вероятность попадания в мишень для первого стрелка равна 0,6, для второго — 0,5, а для третьего — 0,4.

Ответ: $\frac{10}{19}$.

3.11. Производится отбор экспертов из десяти человек, включающих трех экспертов высшей квалификации, четырех — первой категории, двух — второй категории и одного — третьей категории. Имеется двадцать контрольных вопросов. Эксперт высшей квалификации может ответить на все вопросы, первой категории — на 16 вопросов, второй — на 10 вопросов и третьей — на 5. Вызванный наугад эксперт ответил на 3 произвольно заданных вопроса. Какова вероятность, что этот эксперт из высшей категории?

Ответ: $\approx 0{,}58$.

4. Независимые испытания и схема Бернулли. Понятие о цепях Маркова

4.1. Последовательность испытаний

Согласно определению 2 п. 3.3 событие A не зависит от события B, если вероятность события A не меняется от появления или непоявления B.

Определение 1. Осуществление комплекса условий S на практике называют *испытанием* (экспериментом или опытом). Неоднократное повторение испытаний называют *последовательностью испытаний*.

Определение 2. Испытания $E_1, E_2, \ldots, E_n$ называются *независимыми*, если вероятность того или иного исхода каждого испытания не зависит от исходов других испытаний.

Замечание 1. Независимые испытания могут производиться как при одинаковом комплексе условий S, так и при различных условиях S_i. В первом случае вероятность наступления некоторого события A во всех испытаниях одна и та же, во втором — меняется от испытания к испытанию.

Определение 3. Пусть имеется $\{E_n\}$ — последовательность независимых испытаний и A есть некоторое событие, тогда *успехом* называют каждое наступление события A в этой последовательтельности испытаний и *неуспехом* — наступление противоположного события $\bar{A}$.

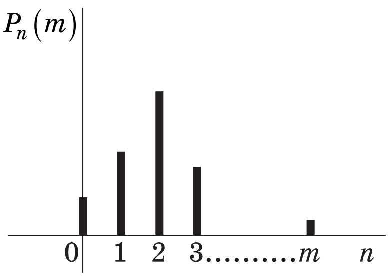
Схема Бернулли. Пусть p — вероятность успеха (A), а $q=(1-p)$ — вероятность неуспеха; пусть также $P_n(m)$ — вероятность того, что в серии n испытаний будет ровно m успехов. Ставится задача: определить $P_n(m)$. Имеем:

$$\underbrace{A, A, \ldots, A}_{m \text{ раз}}, \underbrace{\bar{A}, \bar{A}, \ldots, \bar{A}}_{(n-m) \text{ раз}}$$

$$\underbrace{\overbrace{p, p, \ldots, p}\ \overbrace{q, q, \ldots, q}}_{p^m \cdot q^{n-m}}$$

(в силу независимости событий)

Другие комбинации серии, очевидно, будут иметь такую же вероятность. Всего комбинаций будет C_n^m. Тогда в силу несовместности комбинаций общая вероятность: $\boxed{P_n(m)=C_n^m p^m q^{n-m}}$



$$P_n(m\geq)=\sum_{k=m}^{n} P_n(k)=1-\sum_{k=0}^{m-1} P_n(k)$$

$P_n(1)=1-q^n$ — хотя бы 1 раз

4.2. Наивероятнейшее число успехов в серии из n испытаний

Утверждение 1. В серии из n независимых испытаний существует наивероятнейшее число успехов m_0 события A, определяемое из неравенства, т. е.

$$\left.\begin{array}{l} n \text{ — независимых испытаний} \\ A \text{ — событие, p — вер. успеха} \\ q \text{ — вер. неуспеха} \end{array}\right\} \Rightarrow \begin{array}{l} \text{существует наивероятнейшее} \\ \text{число успехов} \\ m_0\colon np-q \le m_0 \le np+p. \end{array}$$

Доказательство:

$$\frac{P_n(m)}{P_n(m+1)} = \frac{C_n^m \cdot p^m \cdot q^{n-m}}{C_n^{m+1} \cdot p^{m+1} \cdot q^{n-m-1}} = \frac{n!(m+1)!(n-m-1)!}{m!(n-m)!n!} \cdot \frac{q}{p} =$$

$$= \frac{m+1}{n-m} \cdot \frac{q}{p} + 1 - 1 = 1 + \frac{mq+q-np+mp}{(n-m)p} = 1 + \frac{m-(np-q)}{(n-m)p}.$$

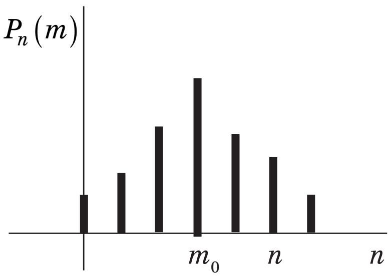
Случаи:

1) $m < (np-q) \Rightarrow P_n(m) < P_n(m+1)$

2) $m > (np-q) \leftrightarrow m+1 > np+p \Rightarrow P_n(m) > P_n(m+1)$

Тогда наивероятнейшее число m_0 должно соответствовать $\frac{P_n(m)}{P_n(m+1)} \to 1$ и, следовательно,

$np-q \le m_0 \le np+p \Rightarrow p - \frac{q}{n} \le \frac{m_0}{n} \le p + \frac{p}{n} \therefore \lim\limits_{n\to\infty} \frac{m_0}{n} = p$, где $\frac{q}{n}$, $\frac{p}{n} \to 0$ при $n \to +\infty$.



4.3. Предельные теоремы схемы Бернулли

Теорема 1 (Я. Бернулли).

$$\left.\begin{matrix} n \text{ — независимых испытаний} \\ m \text{ — число успехов события } A \end{matrix}\right\} \Rightarrow \forall \varepsilon > 0. \ \lim_{n\to\infty} P\left\{\left|\frac{m}{n} - p\right| > \varepsilon\right\} = 0,$$

Другими словами, при неограниченном увеличении числа независимых испытаний вероятность уклонения отношения числа успехов к числу испытаний от вероятности успеха p события A стремится к нулю.

Доказательство (см. литературу).

Теорема 2 (локальная теорема Муавра–Лапласа).

$$\left.\begin{matrix} n \text{ — независимых испытаний} \\ m \text{ — число успехов события } A \\ p, q \text{ — вер. успеха и неуспеха } (A) \\ n \to \infty \end{matrix}\right\} \Rightarrow$$

$$\lim_{n\to\infty} P_n(m) : \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(m-np)^2}{2npq}} =$$

$$= 1 \therefore \mathcal{P}_n(m) \cong \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(m-np)^2}{2npq}} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right)$$

Это значит, что вероятность числа успехов в схеме Бернулли в пределе описывается экспоненциальным дискретным полиномом Гаусса.

Замечание 2.

$a < \dfrac{m-np}{\sqrt{npq}} < b$ при больших n и m, a и b — ограничены,

$$\frac{m-nq}{a < \dfrac{m-np}{\sqrt{npq}} < b} \equiv x \therefore \text{ запись } P_n(m) \text{ упрощается.}$$

Доказательство (см. литературу).

Теорема 3 (интегральная теорема Муавра–Лапласа).

$$\left.\begin{array}{l} n \text{ — независимых испытаний} \\ m \text{ — число успехов события } A \\ p,\ q \text{ — вер. успеха и неуспеха } (A) \\ t_1 < t_2 < +\infty \end{array}\right\} \Rightarrow$$

$$\lim_{n\to\infty} P\left\{ t_1 \le \frac{m-np}{\sqrt{npq}} < t_2 \right\} = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-t^2/2}\,dt.$$

Другими словами, вероятность нахождения нормированного уклонения числа успехов от среднего значения в заданном интервале в пределе определяется через площадь криволинейной трапеции под гауссовой кривой.

Доказательство (см. литературу).

Замечание 3. Если проводится последовательно увеличивающаяся n серия испытаний, в каждой из которых применяется схема Бернулли с вероятностью успеха $p_n \to 0$, то такую схему называют схемой Пуассона. Если $p_n = \dfrac{\lambda}{n}$, то имеет место

$P_n(m) \underset{n\to\infty}{\to} \dfrac{\lambda^m}{m!} e^{-\lambda}$ — теорема Пуассона (закон редких событий).

4.4. Простая и однородная цепи Маркова

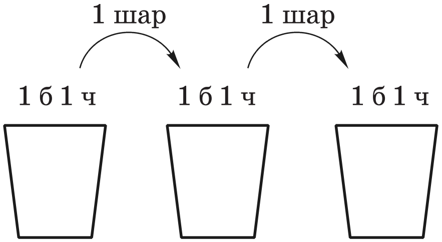
Цепи Маркова являются довольно удобными математическими моделями для описания и исследования систем.

Пусть $\{A_1, A_2, \ldots, A_k\}$ — полная группа событий и производится последовательность испытаний. В каждом отдельном испытании появляется одно из этих событий.

Определение 4. Последовательность испытаний образует *простую* и *однородную цепь Маркова*, если условная вероятность произойти в $(s+1)$–м испытании событию A_i зависит только от того, какое событие произошло в предыдущем испытании, т. е.

$$P_{A_{i_1}A_{i_2}\ldots A_{i_s}}(A_i)=P_{A_{i_s}}(A_i).$$

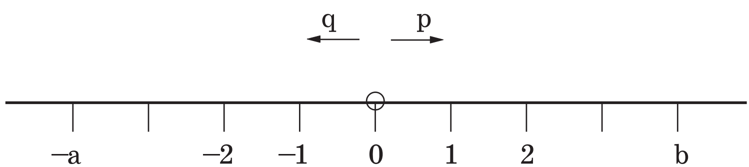
Пример 1. Урновая схема с перекладыванием шаров.



Можно рассматривать другую схему. Пусть система S находится в одном из состояний E_1, E_2,$\geq$, E_k. В моменты времени t_1, t_2, $\geq$ t_s, t_{s+1} система может перейти из одного состояния в другое. Тогда дается другое, эквивалентное определение.

Определение 4а. Последовательность изменения состояния системы S образует *простую* и *однородную цепь Маркова*, если условная вероятность того, что в момент времени t_{s+1} система окажется в состоянии E_i, зависит только от того, в каком состоянии система находилась в момент t_s.

Пример 2. Схема случайных блужданий с поглощающими барьерами. Система имеет нулевое состояние, если она находится в центре «0». Вероятность передвижения за единицу времени влево есть p, а вправо q.



Для задания конечной цепи Маркова с числом состояний E_1, E_2, $\geq$, E_k необходимо иметь:

1) распределение вероятностей P_1, P_2, $\geq$, P_k; P_i — вероятность того, что в начале рассмотрения система находится в состоянии E_i;

2) матрицу переходов вероятностей за один шаг

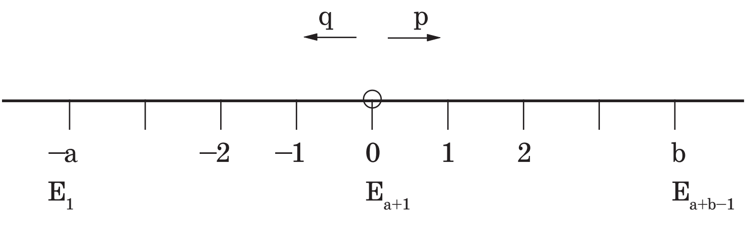
$$\Pi=\begin{pmatrix} P_{11}P_{12}\cdots P_{1k}\\ P_{21}P_{22}\cdots P_{2k}\\ \dots\dots\dots\dots\dots\\ P_{k1}P_{k2}\cdots P_{kk}\end{pmatrix},$$

где P_{ij} — вероятность перехода системы из i-го состояния в j-е за один шаг (i — «система была», j — «система будет»).

Определение 5. Матрицу переходов Π называют *стохастической*, если ее элементы неотрицательны и все построчные суммы равны единице, т. е. $P_{ij}\geq 0, \sum_{j=0}^{k} P_{ij}=1$ для любых $i=1, 2, \geq, k$.

Пример 1. $\Pi_{q/p}=\begin{pmatrix}\frac{2}{3} & \frac{1}{3}\\ \frac{1}{3} & \frac{2}{3}\end{pmatrix}.$

Пример 2. $\Pi=\begin{pmatrix} 1\ 0\ 0\ \dots\ 0\\ q\ 0\ p\ \dots\ 0\\ 0\ q\ 0\ p\dots 0\\ \dots\dots\dots\dots\dots\\ 0\ 0\ 0\ 0\dots 1\end{pmatrix}.$



Всякая стохастическая матрица может задавать цепь Маркова. Пусть задана цепь Маркова, т. е. распределение и матрица переходов. Вероятность перехода из состояния E_i в E_j за n шагов есть $P_{ij}(n)$.

$E_i \xrightarrow{n\ шагов} E_j$. Тогда матрица $\Pi_n = \begin{pmatrix} P_{11}(n) & P_{12}(n) & \cdots P_{1k}(n) \\ P_{21}(n) & P_{22}(n) & \cdots P_{2k}(n) \\ \ldots & \ldots & \ldots \\ P_{k1}(n) & P_{k2}(n) & \cdots P_{kk}(n) \end{pmatrix}$.

Можно рассмотреть переход системы из состояния E_i в состояние E_j другим путем по сравнению с вышеуказанным за те же n шагов.

$$\begin{tikzcd} E_i \arrow[rr, "n\ \text{шагов}"] \arrow[dr, "s\ \text{шагов}"'] & & E_j \\ & E_r \arrow[ur, "n-s\ \text{шагов}"'] & \end{tikzcd}$$

Тогда для вероятности перехода путем $i-r-j$ за $s+(n-s)=n$ шагов имеем $P_{ij}(n)=P_{ir}(s)P_{rj}(n-s)$. Поскольку таких путей переходов имеется k, то в силу их несовместности общая вероятность перехода за n шагов определится следующим образом:

$$P_{ij}(n)=\sum_{r=1}^{k} P_{ir}(s)P_{rj}(n-s).$$

Это выражение называют уравнением Маркова. Из этого уравнения следует, что для матрицы перехода справедливо соотношение: $\Pi_n=\Pi_n\Pi_{n-s}$.

Тогда имеем $\Pi_2=\Pi\Pi=\Pi^2$, $\Pi_3=\Pi_2\Pi=\Pi^3$, $\geq$, $\Pi_n=\Pi^n$.

Таким образом, чтобы определить вероятность перехода системы за n шагов, нужно матрицу перехода за один шаг возвести в n-ю степень.

Определение 6. Состояние E_j называется *достижимым* из состояния E_i, если существует целое число $s \geq 1$, что вероятности перехода больше нуля $(P_{ij} > 0)$, в противном случае состояние E_j называется *недостижимым*.

Определение 7. Состояния $E_{i_1}, E_{i_2}, \ldots, E_{i_k}$ образуют *замкнутый класс*, если все остальные состояния системы недостижимы из состояний $E_{i_1}, E_{i_2}, \ldots, E_{i_k}$.

Определение 8. Если цепь не имеет замкнутых классов состояний, то такая цепь называется *неприводимой.*

Пусть вероятность того, что система вернется в состояние E_i впервые ровно через n шагов есть $P_{ii}(n)$, тогда вероятность того, что система вернется в исходное состояние есть $P_i = \sum_{n=1}^{\infty} P_{ii}(n)$.

Определение 9. Если вероятность возврата системы в исходное состояние равна единице (P_i=1), то состояние E_i называют *возвратным состоянием,* если меньше единицы — то *невозвратным.*

Определение 10. Состояние E_i называют *периодическим,* если система возвращается в состояние через число шагов, кратное периоду T, т. е. при $P_{ii}(kT) > 0$ и $P_{ii}(n) = 0$ при $n \neq kT$, в противном случае состояние E_i называется *непериодическим.*

Определение 11. Цепь называется *транзитивной*, если система может перейти в любое состояние для выбранного числа шагов, т. е. $\exists\ s \geq 1,\ P_{ij}(s) > 0$ для $\forall i, j = 1,\ 2, \ldots,\ k$.

Теорема (без доказательства). Если простая и однородная цепь Маркова транзитивна, то для вероятности перехода из состояния E_i в состояние E_j за n шагов $P_{ij}(n)$ существует предел, равный вероятности P_j, не зависящий от начала, т. е. $\lim\limits_{n \to \infty} P_{ij}(n) = P_j$.

По существу, эта теорема утверждает, что настоящее не зависит от «глубокого» прошлого, или что зависимость настоящего от прошлого с течением времени ослабевает.

4.5. Примеры и задачи

Примеры

Пример 4.1. Из таблицы случайных чисел наудачу выписаны 200 двузначных чисел (от 0 до 99). Определить вероятность того, что среди них число 33 встречается три раза.

Решение:

$$P_n(m)=C_n^m p^m q^{n-m};\ n=200;\ m=3;\ p=\frac{1}{100}=0{,}01;\ q=0{,}99\Rightarrow$$

$$P_n(m)=C_{200}^3\cdot 0{,}01^3\cdot 0{,}99^{197}\approx 0{,}18.$$

Пример 4.2. В системе контроля окружающей среды в заданном районе производится пять независимых заборов почвы с вероятностью обнаружения отравляющих веществ 0,6. Для принятия решения об объявлении района опасной зоной требуется не менее трех обнаружений. Какова вероятность, что район будет объявлен опасной зоной?

Решение:

$$P_n(\geq m)=1-\sum_{k=0}^{m-1}P_n(k).$$

$$m_5(\geq 3)=1-C_5^0\,\mathrm{p}^0 q^5-C_5^1 p\cdot q^4-C_5^2 p^2 q^3=1-1\cdot 0{,}6^0\cdot 0{,}4^5-$$

$$-5\cdot 0{,}6\cdot 0{,}4^4-10\cdot 0{,}6^2\cdot 0{,}4^3\cong 0{,}761.$$

Пример 4.3. При передаче информации по заданному каналу связи вероятность искажения каждого донесения равна 0,02. Всего передано четыре донесения. Какова вероятность того, что среди переданных донесений будет не более одного искажения?

Решение:

$$P_n(\le m) = \sum_{k=0}^{m-1} P_n(k).$$

$$P_4(\le 1) = C_4^0 \cdot 0{,}02^0 \cdot 0{,}98^4 + C_4^1 \cdot 0{,}02 \cdot 0{,}98^3 =$$

$$= 1 \cdot 1 \cdot 0{,}98^4 + 4 \cdot 0{,}02 \cdot 0{,}98^3 \cong 0{,}995.$$

Пример 4.4. Самолет МЧС проводит операцию по тушению лесного пожара. Сброшено пять водяных контейнеров с вероятностью попадания каждого в зону огня 0,7. Определить наивероятнейшее число контейнеров, попавших непосредственно в зону огня.

Решение:

$np - q \le m_0 \le np + p$; $n = 5$; $p = 0{,}7$; $q = 1 - 0{,}7 = 0{,}3$;

$5 \cdot 0{,}7 - 0{,}3 \le m_0 \le 3{,}5 + 0{,}7 \Rightarrow m_0 = 4$.

Задачи

4.1. В библиотеке имеются книги только по гуманитарным и естественно-научным дисциплинам. Вероятность того, что любой читатель возьмет книгу по гуманитарным и естественно-научным дисциплинам равна соответственно 0,7 и 0,3. Определить вероятность того, что пять читателей подряд возьмут книги или только гуманитарные или только естественно-научные, если каждый берет одну книгу.

Ответ: 0,17.

4.2. Что вероятнее — выиграть у равносильного противника: а) три партии из четырех или пять из восьми; б) не менее трех партий из четырех или не менее пяти партий из восьми? Ничейный результат исключается.

Ответ: а) $\frac{1}{4}$ и $\frac{7}{32}$ — вероятнее выиграть три партии из четырех;

б) $\frac{5}{16}$ и $\frac{93}{256}$ — вероятнее выиграть не менее пяти партий из восьми.

4.3. Какова вероятность того, что при 10 бросаниях игральной кости два раза выпадут 6 очков?

Ответ: $\approx 0{,}29$.

4.4. Монету бросают 100 раз. Сколько раз вероятнее всего выпадет при этом герб? Чему приблизительно равна эта вероятность?

Ответ: 50 раз; $P \cong 0{,}0796$.

4.5. В семье десять детей. Считая вероятности рождения равным 0,5, определить вероятность того, что в данной семье: а) пять мальчиков; б) мальчиков не менее трех, но не более восьми.

Ответ: а) $\dfrac{63}{256}$; б) $\dfrac{997}{1024}$.

4.6. Оптовая база снабжает 12 магазинов, от каждого из которых может поступить заявка на очередной день с вероятностью 0,4, независимо от заявок других магазинов. Найти наивероятнейшее число заявок в день.

Ответ: 5 заявок.

5. Случайные величины и их характеристики

5.1. Понятие случайной величины. Функция распределения

Интуитивные (содержательные) представления:

1) Под *случайной величиной* понимают величину, которая может принять то или иное значение, причем неизвестно заранее, какое именно. Другими словами, *случайная величина* — это переменная величина, значение которой есть число, определяемое исходом некоторого эксперимента. *Случайную величину* можно определить так же, как числовую функцию от элементарного события.

   Случайная величина характеризует количественный результат испытания. Примеры случайных величин:
   а) остаток вклада по выбранному наудачу лицевому счету;
   б) число вызовов на телефонной станции;
   в) продолжительность обслуживания покупателя и т. д.
2) Всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими

им вероятностями называют *законом распределения* случайной величины.
Функцию, выражающую вероятность того, что случайная величина примет значение меньше, чем заданное число x, называют *функцией распределения случайной величины* и обозначают через $F(x)=P\{X<x\}$.

Теоретические (формальные) представления:

Определение 1. Пусть $\{\Omega, F, P\}$ — вероятностное пространство, где $\Omega=\{\omega\}$ — пространство элементарных событий как точечное множество, и пусть $X=f(\omega)$ есть некоторая числовая функция от элементарного события, для которой определена вероятность $P\{\omega: f(\omega)<x\}=P\{X<x\}$. Тогда числовая функция $X=f(\omega)$ называется *случайной величиной*, а функция $F(x)=P\{X<x\}$ для $\forall x \in(-\infty, \infty)$ *функцией распределения*.

Замечание 1. В определении 1 множество $\{\omega: f(\omega)<x\} \subset F$ является событием.

Свойства функции распределения:

1) $0 \leq F(x) \leq 1$ (в силу определения функции как вероятности).
2) $x_2>x_1 \Rightarrow F(x_2) \geq F(x_1)$

$$F(x_2)=P\{X<x\}=\underbrace{P\{X<x_1\}}_{F(x_1)}+\underbrace{P\{x_1 \leq X<x_2\}}_{\geq 0} \Rightarrow F(x_2) \geq F(x_1)$$

   и $P\{x_1 \leq X<x_2\}=F(x_2)-F(x_1)$.
3) $F(x)=F(x-0)$ — функция непрерывна слева (без доказательства).
4) $F(-\infty)=0$; $F(+\infty)=1$ — очевидно.

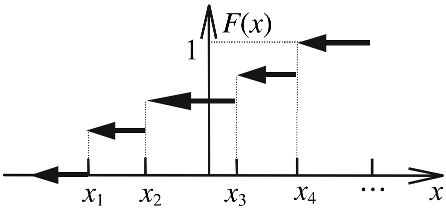
5.2. Дискретные и непрерывные случайные величины

Определение 2. Случайная величина X называется *дискретной*, если она может принимать не более чем счетное число значений.

Дискретная случайная величина обычно задается через функцию распределения или таблицей значений с соответствующими вероятностями $\left|\begin{matrix} X_1 \; X_2 \; X_3 \ldots . X_n \\ P_1 \; P_2 \; P_3 \; \ldots . \; P_n \end{matrix}\right.$, где $\sum_{i=1}^{n} P_i = 1$.

Замечание 2. Если дискретная случайная величина задается законом распределения (таблицей), то функция распределения определяется просто:

$$F(x) = \begin{cases} 0, & X \le x_1 \\ P_1, & x_1 < X \le x_2 \\ & \vdots \\ P_1 + \ldots . + P_k, & x_k \le x \le x_{k+1} \\ 1, & x_n < x < +\infty \end{cases}$$



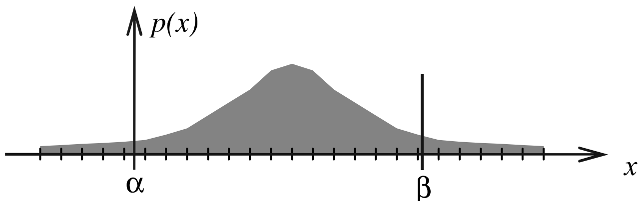
Замечание 3. Если задана функция распределения, то случайная величина определяется с точностью до постоянного множителя.

Определение 3. Случайная величина X называется *непрерывной*, если она может принимать любые значения из некоторого интервала (a, b) и существует неотрицательная функция $p(x) \ge 0$ такая, что вероятность попадания случайной величины в любой интервал определяется через определенный интеграл от этой функции, т. е. $\forall \alpha$ и β: $a \le \alpha < \beta \le b$, $P\{\alpha \le X \le \beta\} = \int_{\alpha}^{\beta} p(x)dx$, причем $\int\limits_{-\infty}^{\infty} p(x)dx = 1.$

Определение 4. Неотрицательная функция $p(x)$ из определения 3 называется *плотностью распределения случайной величины*.

Замечание 4. Очевидно, требование $\int\limits_{-\infty}^{\infty} p(x)dx = 1$ в определении 3 означает $\lim\limits_{x \to \pm\infty} p(x) = 0.$ Тогда геометрическая интерпретация $p(x)$ следующая: площадь под кривой равна единице.

Замечание 5. Непрерывная случайная величина принимает каждое свое значение с вероятностью, равной нулю.



Замечание 6. Для непрерывной случайной величины функция распределения $F(x)$ является дифференцируемой во всех точках промежутка (a, b) и связана с плотностью соотношениями: $F(x) = \int\limits_{-\infty}^{x} p(t)dt;\ F'(x) = p(x),$ т. е. плотность есть производная от функции распределения.

Очевидно также: $P\{\alpha \le X \le \beta\} = F(\beta) - F(\alpha) = \int\limits_{\alpha}^{\beta} p(x)dx.$

5.3. Примеры и задачи

Примеры

Пример 5.1. Из урны, содержащей три белых шара и пять черных шаров, наугад извлекают три шара. Случайная величина m — число вынутых черных шаров. Построить ряд распределения случайной величины m.

Решение: число вынутых черных шаров m может принимать следующие значения: 0, 1, 2, 3. Вероятность того, что среди трех вынутых шаров не будет черных, равна:

$P_0 = \frac{C_3^3}{C_8^3} = \frac{3!\cdot 5!}{8!} = \frac{1}{56}$; один черный шар $P_1 = \frac{C_3^2 \cdot C_5^1}{C_8^3} = \frac{15}{56}$;

два черных $P_2 = \frac{C_3^1 \cdot C_5^2}{C_8^3} = \frac{15}{28}$ и три черных $P_3 = \frac{C_5^3}{C_8^3} = \frac{5}{28}$.

Ряд распределения имеет вид:

m	0	1	2	3
P_m	$^1/_{56}$	$^{15}/_{56}$	$^{15}/_{28}$	$^5/_{28}$

Пример 5.2. Дан закон распределения для случайной величины X.

X_i	–2	0	2	3
P_i	0,1	0,5	0,3	0,1

Составить закон распределения для случайных величин

$$Y_1 = 6-3X,\ Y_2 = X^2.$$

Решение: Возможные значения случайной величины Y_1: -3; 0; 6 и 12. Ряд распределения случайной величины Y_1 имеет вид:

Y_i	–3	0	6	12
P_i	0,1	0,3	0,5	0,1

Ряд распределения случайной величины $Y_2=X^2$ имеет вид:

Y_i	0	4	9
P_i	0,5	0,4	0,1

Задачи

5.1. Экзаменатор задает не более трех дополнительных вопросов студенту. Экзамен считается несданным и заканчивается, как только экзаменатор не получает ответа на очередной вопрос. Вероятность ответа студентом на любой вопрос экзаменатора равна 0,6. Случайной величиной является количество заданных вопросов. Построить ряд распределения случайной величины.

Ответ:

m_i	1	2	3
P_i	0,4	0,24	0,36

5.2. Экзаменационный билет состоит из четырех вопросов. На каждый вопрос даны три возможных ответа, среди которых необходимо выбрать один правильный. Построить ряд распределения числа вопросов, на которые получен правильный ответ. Какова вероятность P того, что методом простого угадывания удастся ответить по крайней мере на три вопроса?

Ответ:

m_i	0	1	2	3	4
P_i	$^{16}/_{81}$	$^{32}/_{81}$	$^{24}/_{81}$	$^{8}/_{81}$	$^{1}/_{81}$

$P={}^{1}/_{9}$.

5.3. Академия закупила 500 стульев. Вероятность того, что при перевозке стульев со склада в академию будет сломан какой-либо стул равна 0,01. Найти вероятность того, что: а) будет сломано ровно 3 стула; б) менее 3 стульев; в) хотя бы один?

Ответ: 0,139; 0,124; 0,04.

5.4. Случайная величина задана плотностью распределения вероятности:

$$p(x)=\begin{cases}0 & x<0\\ A\cdot x\cdot(1-x) & 0<x<1.\\ 0 & x>1\end{cases}$$

Найти функцию распределения и вероятность того, что случайная величина в результате испытания примет значение, заключенное в интервале (0,25; 0,75).

Ответ:

$$F(x)=\begin{cases}0 & x<0\\ x^2\cdot(3-2x) & 0<x<1, P\{0{,}25<x<0{,}75\}=11/16.\\ 1 & x>1\end{cases}$$

5.5. На шоссе установлен контрольный пункт по проверке технического состояния машин. Найти вероятность времени ожидания очередной машины большего чем 10 мин., если время между прохождением машин распределено по показательному закону $p(t)=0{,}2e^{-0{,}2t}$.

Ответ: $P=e^{-2}$.

5.6. Функция распределения случайной величины X (безотказной работы некоторого прибора) равна: $F(x)=1-e^{-x/T}\ (x>0)$.

Найти вероятность безотказной работы прибора за время $x>3T$.

Ответ: 0,05.

6. Числовые характеристики случайной величины. Основные распределения

6.1. Числовые характеристики

Функция распределения (закон распределения для дискретной случайной величины и плотность распределения для непрерывной случайной величины) является исчерпывающей характеристикой случайной величины. Иногда на практике удобно пользоваться числовыми характеристиками случайных величин. Рассматривают две группы характеристик.

I. Характеристики положения

1. Математическое ожидание (среднее, центр распределения).

Определение 1. Пусть $\{\Omega, F, P\}$ — вероятностное пространство и $X = f(\omega)$ — случайная величина. Тогда величина, определяемая по правилу $\int\limits_{\Omega} f(\omega)dP = \int\limits_{-\infty}^{\infty} xdF(x)$ и обозначаемая через EX, называется *математическим ожиданием* случайной величины X, причем $F(x)$ — функция распределения случайной величины X.

Для непрерывной случайной величины $EX = \int\limits_{-\infty}^{+\infty} xp(x)dx$, а для дискретной $EX = \sum\limits_{i=1}^{n} X_i P_i$.

Свойства математического ожидания EX:

1) Математическое ожидание постоянной величины равно самой постоянной, т. е. C — $const \Rightarrow EC = C$.

   В самом деле, $C = \int\limits_{-\infty}^{\infty} CdF(x) = C$.

2) Постоянную можно выносить за знак математического ожидания, т. е. C —$const \Rightarrow ECX = CE\text{X}$.

   В самом деле, $EC = \int\limits_{-\infty}^{\infty} CxdF(x) = CEX$.

3) Математическое ожидание суммы независимых случайных величин равно сумме математических ожиданий этих случайных величин, т. е. X, Y — независимые $\Rightarrow E(X+Y) = EX + EX$.

В самом деле,

$$\int\int_{-\infty}^{\infty}(x+y)dF(x)dF(y)=\int_{-\infty}^{\infty}xdF(x)+\int_{-\infty}^{\infty}ydF(y)=EX+EY.$$

Замечание 1. Определение независимых случайных величин будет дано в разделе 8 «Системы случайных величин». Понятие независимости случайных величин следует из независимости случайных событий.

4) Математическое ожидание произведения независимых случайных величин равно произведению математических ожиданий, т. е. X, Y — независимые $\Rightarrow E(XY)=EX\cdot EY$.

2. Начальный момент k-го порядка.

Определение 2. В условиях определения 1 величина, определяемая по правилу $\alpha_k=\int_{-\infty}^{\infty}x^k dF(x)$, называется *начальным моментом* k-го порядка. Очевидно, $\alpha_k=EX^k$, для непрерывных случайных величин $\alpha_k=\int_{-\infty}^{+\infty}x^k p(x)dx$, а для дискретных $\alpha_k=\sum_i X_i^k P_i$.

Основное свойство: если существует начальный момент k-го порядка, то существуют и все начальные моменты меньшего порядка, т. е. $\exists\alpha_k\Rightarrow\exists\,\alpha_1,\alpha_2,\ldots,\alpha_{k-1}$.

II. Характеристики рассеяния случайной величины

1. Дисперсия

Определение 3. В условиях определения 1 п. 6.1 величина, определяемая по правилу $\int_{-\infty}^{\infty}(x-EX)^2 dF(x)$ и обозначаемая через DX, называется *дисперсией* случайной величины. Для непрерывных случайных величин $DX=\int(x-EX)^2 p(x)dx$, а для дискретных $DX=\sum_i(X_i-EX)^2 P_i$.

Замечание 2. Дисперсия есть математическое ожидание квадрата уклонения от среднего, т. е. $DX = E(X - EX)^2$.

Свойства дисперсии DX:

1) Дисперсия постоянной величины равна нулю: $DC = 0$.
2) Постоянная величина выносится из-под знака дисперсии в квадрате: $D(CX) = C^2 DX$; $D(-X) = DX$.
3) Дисперсия суммы независимых случайных величин равна сумме дисперсий этих величин, т. е. X, Y — независимые $\Rightarrow D(X+Y) = DX + DY$.
4) Сложение случайной величины с постоянной величиной не изменяет дисперсии, т. е. $D(X+C) = DX$.

Определение 4. Случайная величина $y \equiv \dfrac{X - EX}{\sqrt{DX}}$ называется *нормированным уклонением*, а величина $\sqrt{DX}$ — *стандартным уклонением*. $E_y = 0$, а $D_y = 1$.

2. Центральный момент k-го порядка

Определение 5. В условиях определения 1 п. 6.1 величина, определяемая по правилу $\int\limits_{-\infty}^{\infty} (x - EX)^k \, dF(x)$ и обозначаемая через μ_k, называется *центральным моментом* k-го *порядка*.

Другими словами, $\mu_k = E(X - EX)^k$. Очевидно, $\mu_1 = 0$, $\mu_2 = DX$.

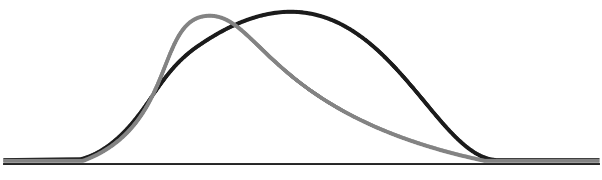
Связь между центральными и начальными моментами

$$\mu_k = \sum_{l=0}^{k} C_k^l (-1)^{k-l} \alpha_1^{k-l} \cdot \alpha_l.$$

3. Асимметрия

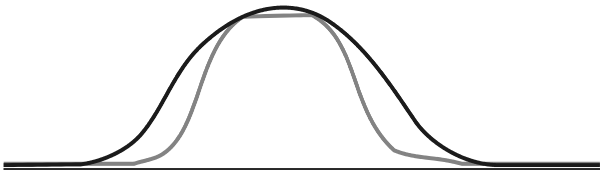
Асимметрия — это величина, определяемая через $A \equiv \dfrac{\mu_3}{\mu_2^{3/2}}$.

Она характеризует «скошенность» распределения.



4. Эксцесс

Эксцесс — это величина, определяемая как $E \equiv \frac{\mu_4}{\mu_2^2} - 3$. Она характеризует «крутизну» распределения.



Замечание 3. В определениях 1—5 характеристики существуют, если интегралы сходятся абсолютно.

5. *Мода Мо* есть значение переменной *x*, соответствующее максимуму плотности распределения для непрерывной случайной величины, т. е. $max\ p(x) = p(Mo)$, и наибольшему значению вероятности для дискретной.
6. *Медиана Ме* есть значение переменной *x*, вероятности быть больше и меньше которой для случайной величины X одинаковы и равны одной второй, т. е.

$$P\{X < Me\} = Bep\{X > Me\} = \frac{1}{2} = F(Me).$$

6.2. Основные дискретные распределения

1. **Вырожденное распределение:**

   а) Закон распределения $P(X = a) = 1$.

   б) Функция распределения $F(x) = \begin{cases} 0, x < a \\ 1, x \geq a \end{cases}$.

   в) Математическое ожидание $EX = p$.

   г) Дисперсия $DX = pq$.

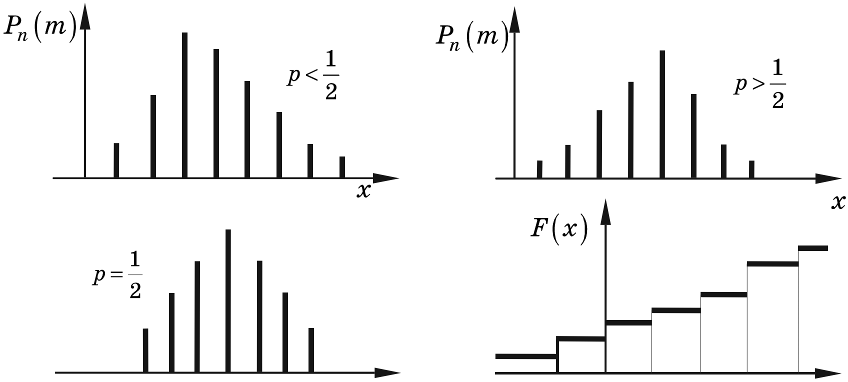
2. **Биномиальное распределение:**

   а) Закон распределения $\underbrace{P(X = m)}_{P_n(m)} = C_n^m p^m q^{n-m}$.

б) Функция распределения $F(x)=\begin{cases} 0,\ x<0 \\ \sum C_n^m p^m q^{n-m},\ x\le n. \\ 1,\ x>n \end{cases}$

в) Математическое ожидание $EX=\sum\limits_{m=1}^{n} mC_n^m p^m q^{n-m}=np.$

г) Дисперсия $DX=npq$.



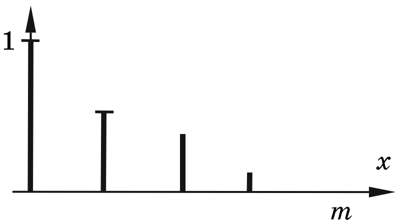
3. **Геометрическое распределение** (вероятность появления события на m-м испытании):

а) Закон распределения $P(x=m)=pq^m$; $m=0,\ 1,\ 2,\ldots$

б) Функция распределения $F(x)=\begin{cases} 0 & ,\quad x<m \\ \sum pq^{m-1} & ,\quad x\ge m<n. \\ 1 & ,\quad x\ge n \end{cases}$

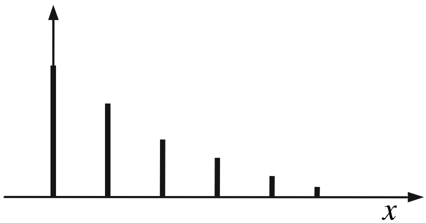
в) Математическое ожидание $EX=\dfrac{1-p}{p}.$

г) Дисперсия $DX = \dfrac{1-p}{p^2}$.



4. Распределение Пуассона:

а) Закон распределения $P\{X=k\} = \dfrac{\lambda^k}{k!} e^{-\lambda}$.



б) Функция распределения $F(x) = \begin{cases} 0, \ x<0 \\ \sum\limits_k \dfrac{\lambda^k}{k!} e^{-\lambda}, \ x \geq 0, \ k<x. \\ 1, \ x \to +\infty \end{cases}$

в) Математическое ожидание

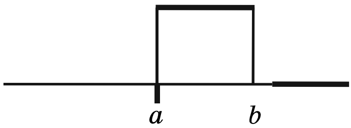
$$EX = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \cdot e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda.$$

г) Дисперсия $DX = \lambda$.

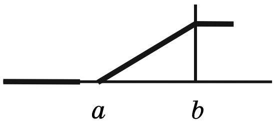
6.3. Непрерывные распределения

1. Равномерное распределение:

а) Плотность распределения $p(x)=\begin{cases}\dfrac{1}{b-a} & ,x\in[a,b]\\ 0 & ,x\notin[a,b]\end{cases}.$



б) Функция распределения $F(x)=\begin{cases}0 & , \quad x\le a\\ \dfrac{x-a}{b-a} & , \quad a\le x<b\\ 1 & , \quad x\ge b\end{cases}.$



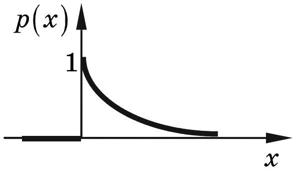
$$P\{\alpha\le X\le\beta\}=\int_{\alpha}^{\beta}\frac{dx}{b-a}=\frac{\beta-\alpha}{b-a}.$$

в) Математическое ожидание $EX=\int_{b}^{a}\dfrac{xdx}{b-a}=\dfrac{a+b}{2}.$

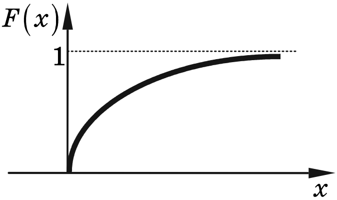
г) Дисперсия $DX=\dfrac{(b-a)^2}{12}.$

2. Показательное распределение (в теории массового обслуживания)**:**
поток событий X — промежуток времени между двумя появлениями событий.

а) Плотность распределения $p(x)=\begin{cases} 0, & x<0 \\ \mu e^{-\mu x}, & x\ge 0 \end{cases}$.



б) Функция распределения $F(x)=1-e^{-\mu x}$.



в) Математическое ожидание $EX=\int\limits_0^{\infty} x\mu e^{-\mu x}dx=\frac{1}{\mu}$.

г) Дисперсия $DX=\frac{1}{\mu^2}$.

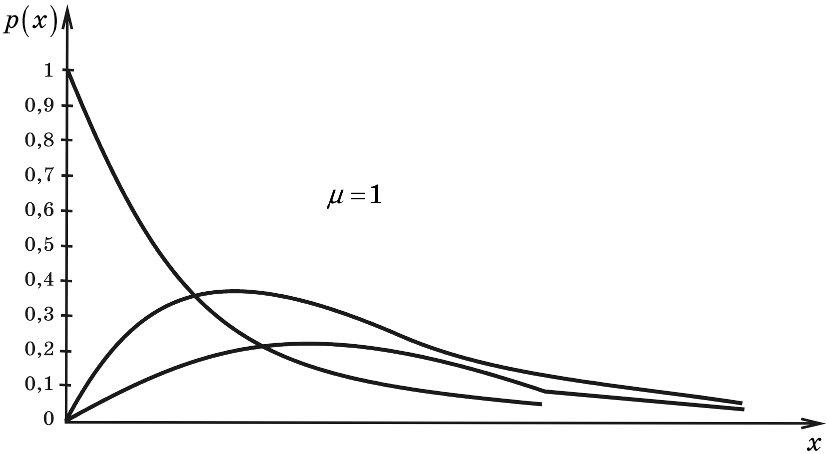
3. **Гамма-распределение** (используется для аппроксимации опытных распределений):
X — время, необходимое для появления заданного числа α событий.
а) Плотность распределения

$$p(x)=\begin{cases} \dfrac{\mu^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1} & ,\ x>0 \\ 0 & ,\ x\le 0 \end{cases},\ \text{где } \Gamma(\alpha)=\int\limits_0^{\infty} x^{\alpha-1}e^{-x}dx.$$

б) Функция распределения $F(x)=\begin{cases} \dfrac{\mu^{\alpha}}{\Gamma(\alpha)}\int\limits_0^{x} u^{\alpha-1}e^{-\mu u}du & ,\ x\ge 0 \\ 0 & ,\ x<0 \end{cases}$.

в) Математическое ожидание $EX = \frac{\alpha}{\mu}$.

г) Дисперсия $DX = \frac{\alpha}{\mu^2}$.

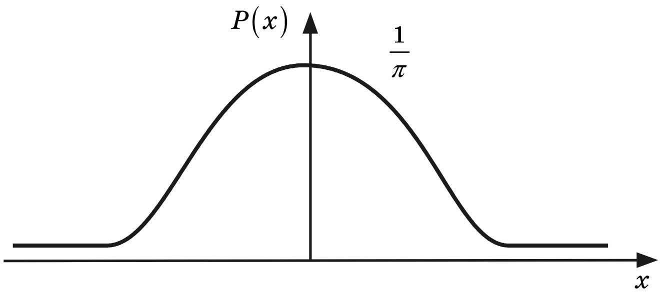


4. Распределение Коши:

а) Плотность распределения $p(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < +\infty$.

б) Функция распределения $F(x) = \frac{1}{2} + \frac{1}{\pi} arctgx$.

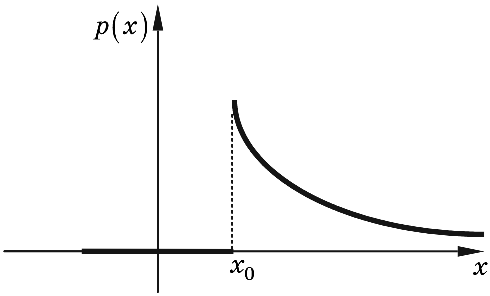
в) Математическое ожидание $EX = 0$.
г) Дисперсия $DX = +\infty$.



5. **Гиперболическое распределение (Парето)** (в экономической статистике):

а) Плотность распределения $p(x)=\begin{cases} 0 & , x \le x_0 \\ \dfrac{\alpha}{x_0}\left(\dfrac{x_0}{x}\right)^{\alpha+1} & x > x_0 \end{cases}$

$\alpha > 0,\ x_0 > 0.$



б) Функция распределения $F(x)=1-\left(\dfrac{x_0}{x}\right)^{\alpha}$ — распределение годовых доходов.

в) Математическое ожидание $EX=\dfrac{\alpha}{\alpha-1}x_0$.

г) Дисперсия

$$DX=\begin{cases} \dfrac{\alpha}{(\alpha-1)(\alpha-2)}x_0^2, & \alpha>2 \\ +\infty & \alpha\le 2 \end{cases} \quad \| \quad \begin{matrix} \alpha\approx 0{,}5 \text{ — распределение Уиллиса} \\ \alpha=1 \text{ — закон Лотки, Ципфа} \end{matrix}$$

$\alpha \in (0,2]$.

6.4. Примеры и задачи

Примеры

Пример 6.1. Дана функция распределения вероятности случайной величины X:

$$F(x)=\begin{cases} 0, & x<-\pi/2 \\ 0{,}5\cdot(1+\sin x), & -\pi/2\le x\le \pi/2. \\ 1, & x>\pi/2 \end{cases}$$

Найти математическое ожидание и дисперсию заданной случайной величины.

Решение:

Плотность вероятности случайной величины равна первой производной от функции распределения:

$$p(x)=F'(x)=\begin{cases} 0, & x<-\pi/2 \\ 0{,}5\cdot\cos x, & -\pi/2\le x\le \pi/2. \\ 0, & x>\pi/2 \end{cases}$$

Математическое ожидание и дисперсия случайной величины равны:

$$EX=\int_{-\pi/2}^{\pi/2} x\cdot 0{,}5\cdot\cos x\,dx=\pi/2$$

$$DX=\int_{-\pi/2}^{\pi/2} (x-EX)^2\cdot 0{,}5\cdot\cos x\,dx=2.$$

Пример 6.2. Отклонение лайнера от генерального курса есть случайная величина, функция распределения которой имеет вид:

$$F(x)=\begin{cases} 0 & x\le -m \\ A+\dfrac{B}{\pi}\arcsin\dfrac{x}{m}, & -m<x<m, \\ 1 & x\ge m \end{cases}$$

где A и B — неизвестные параметры, m — максимальное значение отклонения. Найти значения постоянных A и B, математическое ожидание, дисперсию, моду и медиану случайной величины.

Решение:

1) $$F(-\infty)=F(-m)=0=A+\frac{B}{\pi}\arcsin\left(\frac{-m}{m}\right)=A+\frac{B}{\pi}\cdot\left(-\frac{\pi}{2}\right)=A-\frac{B}{2};$$

$$F(+\infty)=F(m)=1=A+\frac{B}{\pi}\arcsin\left(\frac{m}{m}\right)=A+\frac{B}{\pi}\cdot\frac{\pi}{2}=A+\frac{B}{2};$$

тогда из системы линейных уравнений $\begin{cases} A-\dfrac{B}{2}=0 \\ A+\dfrac{B}{2}=1 \end{cases}$

находим $A=\dfrac{1}{2}; B=1.$

2) $$p(x)=\begin{cases} 0, \; x\notin(-m,\, m) \\ \left(\dfrac{1}{2}+\dfrac{1}{\pi}\arcsin\dfrac{x}{m}\right)'=\dfrac{1}{\pi}\cdot\dfrac{1}{\sqrt{m^2-x^2}}, \; x\in(-m,\, m) \end{cases}$$

$$EX=\int_{-m}^{m}\frac{1}{\pi}\frac{x}{\sqrt{m^2-x^2}}dx=\frac{1}{\pi}\int_{-m}^{m}\frac{xdx}{\sqrt{m^2-x^2}}=0;$$

$$DX=\int_{-m}^{m}\frac{1}{\pi}\cdot\frac{x^2}{\sqrt{m^2-x^2}}dx=\frac{m^2}{2}; \; \max_x p(x)=p(0)\Rightarrow M_0=0;$$

$$\frac{1}{2}=F(M_e);\ \frac{1}{2}+\frac{1}{\pi}\arcsin\frac{M_e}{m}=\frac{1}{2}\Rightarrow M_e=0.$$

Пример 6.3. Конфликты в работе некоторого коллектива за период времени (0, t) есть случайная величина, имеющая функцию распределения $F(t)=\begin{cases}0,\ t\le 0\\ 1-e^{-t/t_0},\ t>0\end{cases}$, где t — среднее время между конфликтами. Найти плотность распределения и вероятность бесконфликтной работы в коллективе ко времени $t=t_0$.

Решение:

1) $p(t)=F'(t)=\begin{cases}0,\ t\le 0\\ \dfrac{1}{t_0}e^{-t/t_0},\ t>0\end{cases};$

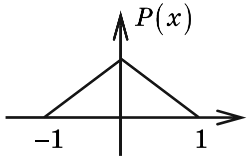
2) $P=1-F(t_0)=1-1-e^{-1}=\dfrac{1}{e}\cong 0{,}368.$

Задачи

6.1. Случайная величина имеет функцию распределения вида $F(x)=A+B\mathrm{arctg}\dfrac{x}{2}$, где A и B — неизвестные параметры. Найти параметры A и B, плотность распределения $p(x)$ и вероятность попадания случайной величины в интервал (−2, 2).

Ответ: $A=\dfrac{1}{2};\ B=\dfrac{1}{\pi};\ p(x)=\dfrac{2}{\pi}\dfrac{1}{x^2+4};\ P=\dfrac{1}{2}.$

6.2. Случайная величина подчинена «закону равнобедренного треугольника» на участке от -1 до $+1$. Написать выражение для плотности распределения, найти математическое ожидание, дисперсию, моду и медиану.

Ответ: а)  $p(x)=\begin{cases}1-|x|, x\in(-1,1)\\ 0, x\notin(-1,1)\end{cases}$;

б) $EX=0$; в) $DX=1/6$; г) $M_0=0$; д) $M_e=0$.

6.3. Вероятность срабатывания таксофона 0,8. Найти математическое ожидание и дисперсию случайной величины — количества опущенных жетонов до срабатывания таксофона.

Ответ: $5/4$ и $5/16$.

6.4. Случайная величина X в интервале (2; 4) задана плотностью распределения вероятности: $p(x)=-\frac{3}{4}x^2+\frac{9}{2}x-6$. Найти моду, медиану и математическое ожидание величины X.

Ответ: $X_{Mo}=3$; $X_{me}=3$; $EX=3$.

7. Семейство нормальных распределений

7.1. Нормальное распределение

Нормальное распределение играет фундаментальную роль в силу центральной предельной теоремы Ляпунова, основной смысл которой сводится к тому, что распределение суммы бесконечно увеличивающегося числа n попарно независимых, с ограниченной дисперсией, произвольно распределенных случайных величин стремится к нормальному закону. Используется запись $X\in N(a,\sigma^2)$. Нормальные случайные величины являются наиболее адекватными моделями для подавляющего числа наблюдаемых величин на практике. Поэтому нормальное распределение рассматривается подробнее других. Как будет показано, из него можно получить другие.

а) Плотность распределения $p(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$ при $x\in(-\infty,\infty)$.

б) Функция распределения $F(x)=\dfrac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-\frac{(x-a)^2}{2\sigma^2}}dx$.

Основные свойства распределения:

1) Параметры распределения a и σ полностью определяют распределение, причем $\sigma>0$, $a\in(-\infty,\infty)$.
2) $EX=a$; $DX=\sigma^2$. Докажем первое равенство:

$$EX=\int\limits_{-\infty}^{\infty}x\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}xe^{-\frac{(x-a)^2}{2\sigma^2}}dx=\left|\begin{matrix}\dfrac{x-a}{\sigma}\equiv z\\ dx=\sigma dz\end{matrix}\right|=$$

$$=\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}(\sigma z+a)e^{-\frac{z^2}{2}}\cdot\sigma dz=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}(\sigma z+a)e^{-\frac{z^2}{2}}dz=$$

$$=\frac{\sigma}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}z\cdot e^{-\frac{z^2}{2}}dz+\frac{a}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}e^{-\frac{z^2}{2}}dz=0+\frac{a\sqrt{2\pi}}{\sqrt{2\pi}}=a.$$

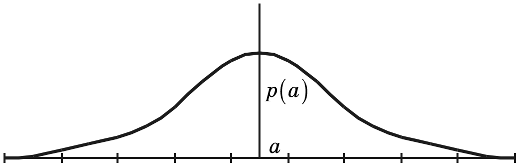
a — среднее (центральная точка). Аналогично доказывается $DX=\sigma^2$.

3) $p(x)$ имеет единственный max при $x=a\therefore p(a)=\dfrac{1}{\sigma\sqrt{2\pi}}$.

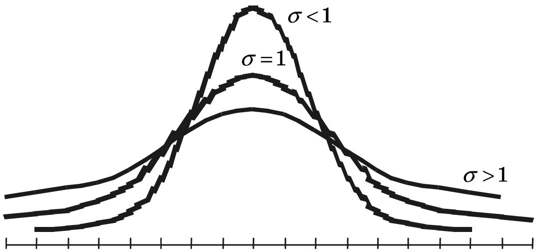
   График $p(x)$ симметричен относительно прямой $x=a$. Около нее группируются значения случайной величины.
4) График $p(x)$ не пересекает ось x, при $x\to\pm\infty$ кривая асимптотически приближается к оси абсцисс.
5) Площадь под кривой $p(x)$ равна 1.

$$P\{|x-a|\le1{,}965\sigma\}=0{,}95$$

$$P\{|x-a|\le3\sigma\}=0{,}997$$



6) σ характеризует меру рассеяния от центра.
$\sigma < 1$ — кривая круче, $\sigma > 1$ — кривая более полога по сравнению с $\sigma = 1$.



$A = 0$ и $E = 0$

7) «Устойчивость» распределения:

$$X_i \in N\left(a_i, \sigma_i^2\right) \Rightarrow \sum_1^n X_i \in N\left(a, \sigma^2\right), \text{ где } a = \sum a_i \text{ и } \sigma^2 = \sum \sigma_i^2.$$

7.2. Стандартное нормальное распределение. Функции Гаусса и Лапласа

Если принять замену $\dfrac{x-a}{\sigma} = z$, то функция $p(x)$ может быть представлена как $p\left(\sigma z + a\right) \equiv \varphi\left(z\right) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ и может рассматриваться как плотность нормального распределения с параметрами 0 и 1, которое называют *стандартным нормальным законом распределения*.

В самом деле, функция $\varphi(z)$, называемая *функцией Гаусса,* имеет все свойства плотности:

1) $\varphi(z) \geq 0.$

2) $\int\limits_{-\infty}^{\infty} \varphi(z)dz = 1.$

3) $\varphi(z) = \varphi(-z).$

4) $\lim\limits_{z \to \pm\infty} \varphi(z) = 0.$

Тогда функцию распределения можно записать в виде интеграла с переменным пределом $F(z) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z} e^{-\frac{u^2}{2}} du.$ Для ее определения удобно ввести и табулировать вспомогательную функцию — *функцию Лапласа,* определяемую как $\Phi(z) \equiv \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{z} e^{-\frac{u^2}{2}} du,$ со свойствами:

1) $\Phi(z)$ — монотонно возрастает, т. к.

$$\Phi'(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} > 0 \text{ для} \forall z.$$

2) $\Phi(z)$ — нечетная, т. е. $-\Phi(z) = \Phi(-z)$ и $\Phi(z) = 0.$

3) $\lim\limits_{z \to \infty} \Phi(z) = 0{,}5$ и тогда функция распределения

$$F(z) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{0} e^{-\frac{u^2}{2}} du + \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{z} e^{-\frac{u^2}{2}} du =$$

$$= -\frac{1}{\sqrt{2\pi}} \underbrace{\int\limits_{0}^{-\infty} e^{-\frac{u^2}{2}} du}_{-\frac{\sqrt{2\pi}}{2}} + \Phi(z) = 0{,}5 + \Phi(z).$$

Следовательно, $F(x) = P\{X < x\} = 0{,}5 + \Phi\left(\frac{x-a}{\sigma}\right)$, и тогда вероятность попадания в интервал соответственно будет

$$\boxed{P\{x_1 < X < x_2\} = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right).}$$

При $\begin{matrix} x_1 \equiv a - z\sigma \\ x_2 \equiv a + z\sigma \end{matrix}$ имеем $\boxed{P\{|x-a| < z\cdot\sigma\} = 2\Phi(z).}$

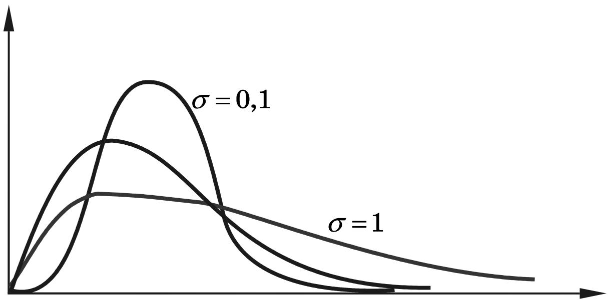
7.3. Логарифмически нормальное распределение

Это распределение используется для описания распределения доходов, банковских вкладов, месячной зарплаты, посевных площадей и т. п. В его основе лежит предельная теорема о стремлении произведения n независимых положительных случайных величин при условии их равномерной малости к логарифмически нормальному закону.

а) плотность распределения $p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - a)^2}{2\sigma^2}}$ $x \in (0, \infty)$.

б) функция распределения

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^x e^{-\frac{(\ln u - a)^2}{2\sigma^2}} \frac{du}{u} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\ln x} e^{-\frac{(v-a)^2}{2\sigma^2}} dv = F_N(\ln x, a, \sigma).$$



Замечание. С семейством нормальных распределений связаны χ^2-распределение (Пирсона), t-распределение (Стьюдента), F-распределение (Фишера-Снедекора), которые будут рассмотрены в разделе статистики.

7.4. Примеры и задачи

Примеры

Пример 7.1. Случайная величина распределена по нормальному закону $X \in N(0,1)$. Вероятность какого события больше: $|X| \le 0,7$ или $|X| > 0,7$?

Решение: вероятности событий определяются с помощью функции Лапласа.

$$P\{|X| \le 0,7\} = 2\Phi(0,7) = 2 \cdot 0,258 = 0,516;$$

$$P\{|X| > 0,7\} = 1 - P\{|X| \le 0,7\} = 1 - 0,516 = 0,484$$

$$\Rightarrow P\{|X| \le 0,7\} > P\{|X| < 0,7\}.$$

Пример 7.2. Случайная величина распределена по нормальному закону $X \in N(0, \sigma^2)$. Вероятность попадания случайной величины в интервал $(-1, 1)$ равна 0,7. Найти среднеквадратическое (стандартное) отклонение σ и записать выражение для плотности.

Решение: $P\{|X| < 1\} = P\left\{|X| < \frac{1}{\sigma} \cdot \sigma\right\} = 2\Phi\left(\frac{1}{\sigma}\right) = 0,35.$

По таблице функций $\Phi(z)$ имеем $\frac{1}{\sigma} \approx 1,04 \Rightarrow \sigma \cong 0,961.$

$$p(x) = \frac{1,04}{\sqrt{2\pi}} e^{-\frac{x^2}{1,827}}.$$

Пример 7.3. Случайная величина $X \in N(a, \sigma^2)$. Необходимо аппроксимировать нормальное распределение равномерным на участке $[\alpha, \beta]$, сохранив неизменным математическое ожидание и дисперсию. Найти α, β и записать распределение.

Решение: Условия равенства средних: $a = \dfrac{\alpha+\beta}{2}$; равенство дисперсий: $\sigma^2 = \dfrac{(\beta-\alpha)^2}{12}$. Решая уравнение относительно α и β, имеем: $\alpha = a - \sigma\sqrt{3}$; $\beta = a + \sigma\sqrt{3}$. Плотность распределения тогда

$$p(x) = \begin{cases} 0, & x \notin \left(a-\sigma\sqrt{3},\ a+\sigma\sqrt{3}\right) \\ \dfrac{1}{2\sigma\sqrt{3}}, & x \in \left(a-\sigma\sqrt{3},\ a+\sigma\sqrt{3}\right) \end{cases}.$$

Задачи

7.1. Математическое ожидание и среднеквадратическое отклонение нормально распределенной случайной величины X соответственно равны 168 и 6. Найти вероятность того, что в результате испытания X примет значение, заключенное в интервале (174; 180).

Ответ: $P\{\alpha < x < \beta\} = 0{,}136$. Значения функции Лапласа находим по таблице.

7.2. Случайная величина $X \in N(a, \sigma)$. Определить абсциссы x_1, x_2 и ординату точек перегиба плотности распределения $p(x)$.

Ответ: $x_1 = a - \sigma$; $x_2 = a + \sigma$; $y \cong \dfrac{0{,}24}{\sigma}$.

7.3. Футболист наносит удар по мячу, прицеливаясь в середину ворот. Ширина ворот 10 м, среднеквадратическое отклонение точки попадания составляет 8 м. Какова вероятность, что мяч попадет в створ ворот?

Ответ: P=0,468.

7.4. Случайная величина $X \in N(a, \sigma)$. Найти:

а) $P\{a-2\sigma < X < a+2\sigma\}$; б) $P\{a-2{,}5\sigma < X < a+2{,}5\sigma\}$;

в) $P\{a-3\sigma < X < a+3\sigma\}$.

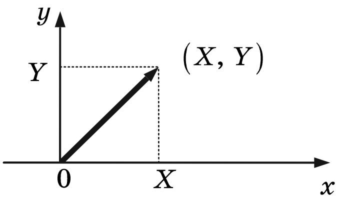
Ответ: а) 0,954; б) 0,987; в) 0,997.

7.5. Имеются две случайные величины $X_1 \in N(0, \sigma_1)$ и $X \in N_2(0, \sigma_2)$, $\sigma_1 > \sigma_2$. Доказать, что для любых $t > 0$ $P\{|X_1| < t\} \le P\{|X_2| < t\}$.

8. Системы случайных величин (случайные векторы)

8.1. Основные понятия о системе случайных величин

Определение 1. Совокупность двух случайных величин $\{X, Y\}$, определенных на одном и том же вероятностном пространстве $\{\Omega, F, P\}$ и рассматриваемых совместно, называется *системой двух случайных величин*. Геометрически это есть *случайная точка* на плоскости xy с координатами X и Y.



Эквивалентные названия системы двух случайных величин: случайный вектор, *двумерная случайная величина*.

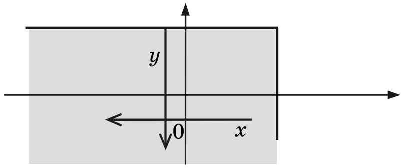
Определение 2. Совокупность трех случайных величин $\{X, Y, Z\}$, рассматриваемых совместно, называется *системой трех случайных величин* или *случайной точкой (случайным вектором)* в трехмерном пространстве.

Аналогично система n случайных величин $\{X_1, X_2, \ldots, X_n\}$ есть n-мерная случайная величина (n-мерный вектор).

Двумерную случайную величину (случайный вектор, систему двух случайных величин) характеризует функция распределения $F(x,y)$.

Определение 3. *Функцией распределения* $F(x,y)$ системы двух случайных величин $\{X, Y\}$ называется вероятность *совместного* выполнения двух событий в виде неравенств $X < x$ и $Y < y$, т. е. $F(x,y) = P\{\omega : f(\omega) = X < x,\ g(\omega) = Y < y\} = P\{X < x,\ Y < y\}$.

Замечание 1. Очевидно, функция распределения $F(x,y)$ есть функция двух переменных. Она характеризует вероятность попадания точки в левую нижнюю область плоскости xy, ограниченную сверху y, а справа $-x$.

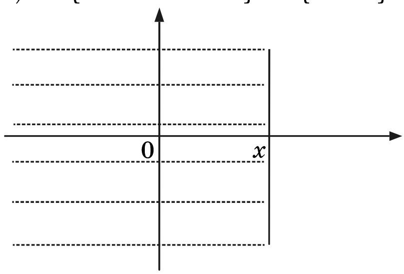


8.2. Свойства функции распределения (системы двух случайных величин)

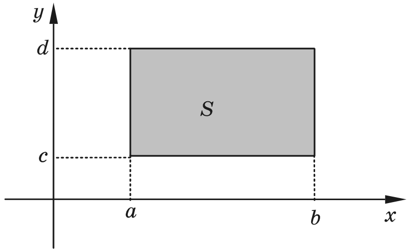
1) $0 \le F(x,y) \le 1$ (в силу свойства вероятности).
2) $F(x,y)$ — неубывающая функция по x и y (из определения функции распределения).
3) $F(x,y)$ —непрерывна слева по x и по y (без доказательства).
4) $F(-\infty,-\infty) = F(-\infty,y) = F(x,-\infty) = 0$ (очевидно).
5) $F(+\infty,y) = F_2(y);\ F(x,+\infty) = F_1(x);\ F(+\infty,+\infty) = 1.$

   В самом деле, пусть

$$F(x,+\infty) = P\{X < x,\ Y < +\infty\} = P\{X < x\} = F_1(x).$$



$$P\{(X,Y) \subset S\} = F(b,d) - F(a,d) - F(b,c) + F(a,c).$$



Замечание 2. Совместная функция распределения полностью определяет одномерные распределения случайных величин X и Y.

8.3. Система двух непрерывных случайных величин (непрерывная двумерная случайная величина)

Определение 4. Двумерная случайная величина называется *непрерывной*, если она принимает любое значение из некоторой области D и существует функция $p(x,y)\geq 0$: такая, что

1) для $\forall S\subset D\ P\{(X,Y)\in S\}=\int\limits_a^b\left(\int\limits_c^d p(x,y)dy\right)dx$, где $S=[a,b]\times[c,d]$

и

2) $\int\limits_{-\infty}^{\infty}\left(\int\limits_{-\infty}^{\infty} p(x,y)dx\right)dy=1.$

Определение 5. Функция $p(x,y)$ из определения 4 называется *плотностью распределения* системы двух случайных величин или двумерной (совместной) плотностью распределения случайных величин X и Y. Поверхность (график функции $p(x,y)$) называют *поверхностью распределения*.

Некоторые дополнительные свойства $p(x,y)$:

1) $p(x,y)=\dfrac{\partial^2 F(x,y)}{\partial x\partial y}$

2) $\int\limits_{-\infty}^{\infty} p(x,y)dy=p_1(x)$ и $\int\limits_{-\infty}^{\infty} p(x,y)dx=p_2(y)$.

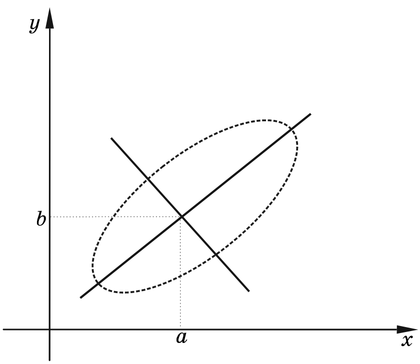
Это следует из того, что $F(x,y)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y}p(t,s)dtds$ и

$$F_1(x)=\int\limits_{-\infty}^{x}\left(\int\limits_{-\infty}^{\infty}p(x_1y)dx\right)dy \therefore F_1'(x)=\int\limits_{-\infty}^{\infty}p(x,y)dy=p_1(x).$$

Пример 1. Система двух нормально распределенных случайных величин нормальна, если

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_1^2}-2r\frac{(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\}$$

$\sigma_1>0$; $\sigma_2>0$ и $|r|<1$. Точка (a,b) — центральная точка (центр группирования), r — параметр, определяющий положение осей эллипса относительно осей координат.



Замечание 3. Соответствующим поворотом осей координат можно сделать их параллельными осям эллипса, что эквивалентно $r=0$ и «разделению» плотностей случайных величин X и Y на произведение $p(x)\cdot p(y)$.

8.4. Примеры и задачи

Примеры

Пример 8.1. Два стрелка независимо один от другого производят по одному выстрелу, каждый по своей мишени. Случайная величина X — число попаданий первого стрелка; Y — второго стрелка. Вероятность попадания в мишень для первого стрелка P_1, для второго — P_2. Построить функцию распределения $F(x,y)$ системы случайных величин (X, Y).

Решение: Так как случайные величины X и Y независимы, то $F(x, y) = P(X<x)\,P(Y<y) = F_1(x)\,F_2(y)$.

Построим функцию распределения $F_1(x)$:

$$F_1(x)=\begin{cases}0, & x\le 0\\ q_1, & 0<x\le 1\\ 1, & x>1\end{cases}, \text{ где } q_1 = 1-P_1.$$

Аналогично

$$F_2(y)=\begin{cases}0, & y\le 0\\ q_2=1-P_2, & 0<y\le 1.\\ 1, & y>1\end{cases}$$

Значения функции $F(x, y)$ даны в таблице:

y	x		
	$x\le 0$	$0<x\le 1$	$1<x$
$y\le 0$	0	0	0
$0<y\le 1$	0	$q_1\times q_2$	q_2
$1<y$	0	q_1	1

Пример 8.2. Пусть плотность распределения системы случайных величин $(X, У)$:

$$p(x,y)=\begin{cases}a\sin(x+y), & 0\le x\le\frac{\pi}{2},\ 0\le y\le\frac{\pi}{2}\\ 0, & x<0,\ y<0,\ x>\frac{\pi}{2},\ y>\frac{\pi}{2}\end{cases}.$$

Требуется: найти коэффициент a; определить функцию распределения системы.

Решение: На основании свойства плотности распределения имеем: $\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}} a\sin(x+y)\,dx\,dy = 1$, откуда: $\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}} a\sin(x+y)\,dx\,dy =$

$$= a\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}} \sin(x+y)\,dx\,dy = a\int_0^{\frac{\pi}{2}} \left(-\cos(x+y)\right)\Big/_0^{\frac{\pi}{2}}\,dx =$$

$$= a\int_0^{\frac{\pi}{2}} \left(-\cos\left(x+\frac{\pi}{2}\right)+\cos x\right)dx = a\left(\int_0^{\frac{\pi}{2}}\left(\cos x+\frac{\pi}{2}\right)dx + \int_0^{\frac{\pi}{2}}\cos x\,dx\right) =$$

$$= a\left(-\sin\left(x+\frac{\pi}{2}\right)\Big/_0^{\frac{\pi}{2}} + \sin x\Big/_0^{\frac{\pi}{2}}\right) = 2a = 1$$

$$a = \frac{1}{2}.$$

Следовательно, плотность распределения двумерной случайной величины (X, Y):

$$p(x,y) = \begin{cases} \frac{1}{2}\sin(x+y), & 0 \le x \le \frac{\pi}{2},\ 0 \le y \le \frac{\pi}{2}; \\ 0, & x<0,\ y<0,\ x>\frac{\pi}{2},\ y>\frac{\pi}{2}. \end{cases}$$

Функция распределения выражается через плотность распределения:

$$F(x,y)=P(X<x,\ Y<y)=\frac{1}{2}\int_0^x\int_0^y \sin(x+y)dxdy=$$

$$=\frac{1}{2}\int_0^x\left(-\cos(x+y)\right)/_0^y\,dx=\frac{1}{2}\int_0^x\left(\left(-\cos(x+y)+\cos x\right)dx=\right.$$

$$=\frac{1}{2}\left(\sin x+\sin y-\sin(x+y)\right).$$

Задачи

8.1. Координаты X и Y случайной точки распределены равномерно внутри прямоугольника, ограниченного абсциссами $x=2$, $x=4$ и ординатами $y=2$, $y=4$. Найти плотность и функцию распределения системы величин X и Y.

Ответ: $p(x,y)=\begin{cases}\frac{1}{4}, & 2\le x\le 4,\ 2\le y\le 4\\ 0, & x<2,\ x>4,\ y<2,\ y>4.\end{cases}$ $F(x, y)=F_1(x)F_2(y)$,

где $F_1(x)=\begin{cases}1, & x>4\\ \frac{x-2}{2}, & 2\le x\le 4,\\ 0, & x<2\end{cases}$ $F_2(y)=\begin{cases}1, & y>4\\ \frac{y-2}{2}, & 2\le y\le 4.\\ 0, & y<2\end{cases}$

8.2. Система случайных величин (X, Y) имеет плотность распределения: $p(x,y)=\dfrac{a}{\pi^2\left(16+x^2\right)\left(25+x^2\right)}$.

Требуется: определить величину a; найти функцию распределения $F(x, y)$.

Ответ: $a=20$, $F(x,y)=\left(\frac{1}{\pi}arctg\frac{x}{4}+\frac{1}{2}\right)\cdot\left(\frac{1}{\pi}arctg\frac{y}{5}+\frac{1}{2}\right)$.

8.3. Система случайных величин (X, Y) распределена с постоянной плотностью внутри квадрата S с абсциссами $x=0$, $x=1$ и ординатами $y=0$, $y=1$. Написать выражение плотности распределения $p(x, y)$. Построить функцию распределения системы.

Ответ: $p(x,y)=\begin{cases}1, & (x,y)\in S\\ 0, & (x,y)\notin S\end{cases}.$

$$F(x,y)=\begin{cases}0, & x\le 0\ \ y\le 0\\ x\cdot y, & 0<x\le 1, 0<y\le 1\\ x, & 0<x\le 1,\ y>1\\ y, & x>1, 0<y\le 1\\ 1, & x>1,\ y>1.\end{cases}.$$

8.4. Бросают две игральные кости. Пусть ξ — сумма очков, выпадающих на их верхних гранях. Написать закон распределения случайной величины ξ.

Ответ:

ξ	2	3	4	5	6	7	8	9	10	11	12
$36P$	1	2	3	4	5	6	5	4	3	2	1

8.5. Законы распределения числа очков, выбиваемых каждым из двух стрелков, таковы:

ξ	1	2	3
P	0,1	0,3	0,6

η	1	2	3
P	0,2	0,3	0,5

Найти закон распределения суммы очков, выбиваемых двумя стрелками.

Ответ:

$\xi+\eta$	2	3	4	5	6
P	0,02	0,09	0,26	0,33	0,30

8.6. Плотность распределения системы случайных величин равна:

$$p(x,y)=\begin{cases} c\left(R-\sqrt{x^2+y^2}\right), & x^2+y^2\le R; \\ 0, & x^2+y^2\le R. \end{cases}$$

Определить: а) постоянную c; б) вероятность попадания в круг радиуса $a<R$ с центром в начале координат.

Ответ: а) $c=\dfrac{3}{\pi R^3}$, б) $p=\dfrac{3a^2}{R^2}\left(1-\dfrac{2a}{3R}\right)$.

8.7. Определить плотность распределения системы двух положительных случайных величин (X, Y) по заданной функции распределения: $F(x,y)=(1-e^{-ax})(1-e^{-by})$, $(x\ge0,\ y\ge0)$.

Ответ: $p(x,y)=abe^{-(ax+by)}$.

8.8. Независимые случайные величины X и Y подчиняются законам равномерной плотности распределения соответственно в интервалах $(-1;\ 1)$ и $(0;\ 2)$. Определить плотность распределения системы (X, Y).

Ответ: $p(x,y)=\begin{cases} \dfrac{1}{4}, -1<x<1,\ 0<y<2; \\ 0,\ x\le-1,\ x\ge1,\ y\le0,\ y\ge2. \end{cases}$

9. Связь случайных величин

9.1. О распределении составляющих случайного вектора

Без ограничения общности рассматривается двумерная случайная величина.

Определение 1. Функции распределения составляющих случайного вектора называются *маргинальными* (частными) распределениями и вычисляются по функциям распределения случайного вектора как (см. п. 8.2 свойство 5) $F_1(x)=F(x,+\infty)$ и $F_2(y)=F(+\infty,y)$.

Определение 2. Функция распределения случайной величины называется *условной функцией распределения*, если она задается как вероятность события совместного выполнения неравенства с равенством $\left\{\begin{matrix} \omega : f(\omega) = X < x \\ \varphi(\omega) = Y = y \end{matrix}\right\} \subset F$,

т. е. $P\{X < x, Y = y\} \equiv F_y(x)$ и $P\{X = x,\ Y < y\} \equiv F_x(y)$.

Другими словами, условная функция распределения есть функция распределения случайной величины при условии, что другая случайная величина приняла определенное значение.

Замечание 1. Условной функции распределения непрерывной случайной величины соответствует *условная плотность распределения* $p_y(x) = \dfrac{\partial F_y(x)}{\partial x}$ и $p_x(y) = \dfrac{\partial F_x(y)}{\partial y}$.

Замечание 2. С учетом определения условной вероятности и свойства 2 п. 8.3 можно записать:

$$\underbrace{p_y(x) = \frac{p(x,y)}{p_2(y)} \quad \text{и} \quad p_x(y) = \frac{p(x,y)}{p_1(x)}}$$

$$\Updownarrow$$

$$p(x,y) = p_y(x)\,p_2(y) = p_x(y)\,p_1(x) = p_1(x)\,p_x(y) = p_2(y)\,p_y(x)$$

(теорема умножения плотностей распределения).

9.2. Независимость и стохастическая зависимость случайных величин

Определение 3. Случайные величины X и Y называются *независимыми*, если совместная функция распределения может быть представлена в виде произведения маргинальных, т. е. $F(x,y) = F_1(x) \cdot F_2(y)$ (переменные разделяются). Это эквивалентно тому, что $P\{X < x,\ Y < y\} = P\{X < x\} \cdot P\{Y < y\}$.

З а м е ч а н и е 3. Для непрерывных случайных величин в определении 3 равенство может быть записано в виде

$$p(x,y)=p_1(x)\cdot p_2(y).$$

З а м е ч а н и е 4. Для дискретных случайных величин условие независимости может быть записано в виде закона распределения как

$$P_{ik}=P\{X=x_i\}\cdot P\{Y=y_k\}=P_i\cdot P_k.$$

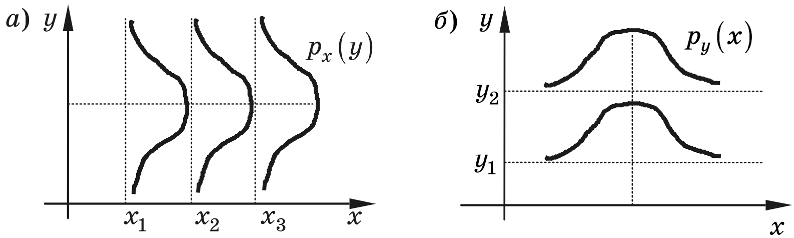
З а м е ч а н и е 5. Условие независимости случайных величин можно выразить по аналогии с независимыми событиями через условные плотности распределения требованием:

$$p_x(y)=p_2(y)\text{ и }p_y(x)=p_1(x).$$

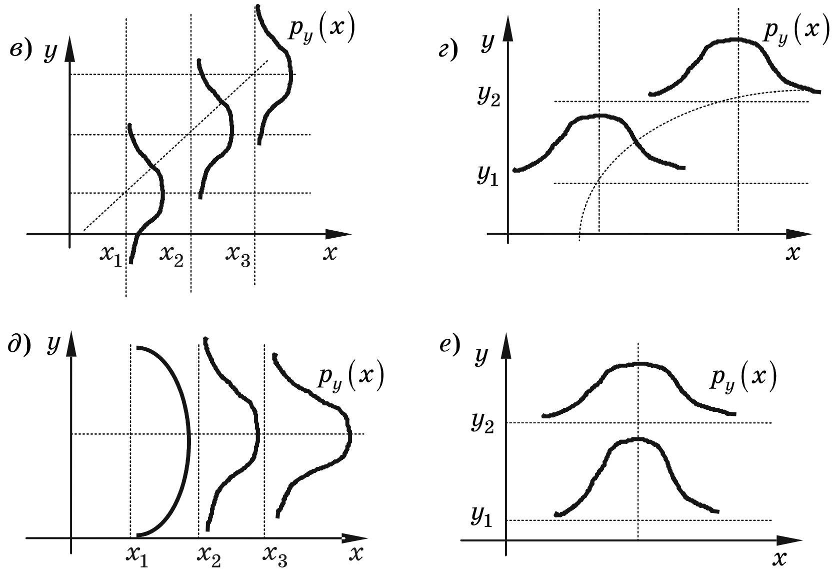
О п р е д е л е н и е 4. Случайные величины X и Y являются *зависимыми*, если $F(x,y)\neq F_1(x)F_2(y)$.

О п р е д е л е н и е 5. Зависимость случайной величины X от Y в виде зависимости через условную плотность распределения $p_y(x)$ или $p_x(y)$ называется *вероятностной* или *стохастической*.

Примеры:



а) и б) — независимые X и Y



в) – е) — стохастическая зависимость.

9.3. Корреляционная зависимость

Характеристикой связи между двумя случайными величинами X и Y могут служить ковариация (корреляционный момент) и коэффициент корреляции. Они являются усредненными, числовыми.

Определение 6. *Ковариацией* случайных величин X и Y называется величина, равная смешанному центральному моменту второго порядка ($\mu_{11}(X,Y)$).

$$cov(X,Y) = \iint\limits_{R_2}(x - EX)(y - EY)\,dF(x,y), \text{ т. е.}$$

$$cov(X,Y) \equiv E(X - EX)(Y - EY) = E(XY) - EX \cdot EY.$$

Определение 7. *Коэффициентом корреляции* двух случайных величин X и Y называется величина (безразмерная), определяемая по правилу: $r(X,Y)=\dfrac{cov(X,Y)}{\sqrt{DX\cdot DY}}$, где DX и DY — конечные дисперсии.

Замечание 6. Коэффициент корреляции характеризует степень «тесноты» линейной зависимости между случайными величинами.

Определение 8. Случайные величины X и Y называются *некоррелированными*, если $cov(X,Y)=r(X,Y)=0$.

Теорема 1 (без доказательства).

Если X и Y — независимые случайные величины, то X и Y — некоррелированные.

Обратное утверждение в общем случае неверно, однако оно справедливо для нормальных распределений.

При нормальном распределении $\left.\begin{matrix} r(X,Y)=0 \\ X,Y\in N \end{matrix}\right\} \Leftrightarrow X,Y$ — независимы.

Свойства коэффициента корреляции:

1) X и Y — независимы $\Rightarrow r(X,Y)=0$.

2) $\left|r(X,Y)\right|\le 1$.

3) $r(X,Y)=\pm 1 \Leftrightarrow Y=aX+b$ (между X и Y существует линейная связь).

9.4. Измерение расстояния между функциями распределения случайных величин (Мера различия случайных величин по функции распределения)

1) Равномерная метрика Колмогорова

$\forall F(x)$ и $G(x)$ $\rho(F,G)=\sup\limits_{x}\left|F(x)-G(x)\right|$.

2) **Метрика Леви** как $inf\,h: L(F,G) = inf\,h \le h.$

$$\forall F(x),\ G(x): \begin{array}{l} F(x-h)-h \le G(x) \le F(x+h)+h, \\ G(x-h)-h \le F(x) \le G(x+h)+h. \end{array}$$

3) **Расстояние по вариации**

Если $\begin{array}{l} P(X\in B) = P\{\omega: X = f(\omega)\in B\} \equiv P(A) \\ P(Y\in B) = P\{\omega: Y = \varphi(\omega)\in B\} \equiv Q(A) \end{array}$

$$Var(P,Q) = \sup_{A\in F}\left|P(A) - Q(A)\right|.$$

Замечание. Иногда для случайной величины X вводится *функция концентрации*, определяемая как

$$Q(x) = \sup_{a} P(a \le X \le a+x) - \infty < a < +\infty.$$

Функция концентрации однозначно определяется функцией распределения, но не наоборот.

9.5. Примеры и задачи

Примеры

Пример 9.1. Система двух случайных величин (X, Y) подчинена закону равномерной плотности распределения внутри круга радиуса 1 с центром в начале координат. Написать выражения для плотностей распределения системы и отдельных случайных величин, входящих в систему. Установить, являются ли случайные величины X и Y зависимыми. В случае их зависимости определить, являются ли они коррелированными.

Решение: Плотность вероятности системы случайных величин (X, Y), равномерно распределенной внутри круга радиуса 1, выражается формулой:

$$p(x,y) = \begin{cases} \dfrac{1}{\pi},\ x^2+y^2 \le 1, \\ 0,\ x^2+y^2 > 1. \end{cases}$$

Тогда плотность распределения отдельных величин, входящих в систему (X, Y):

$$p_1(x,y)=\int_{-\infty}^{\infty}p(x,y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy=\frac{2\sqrt{1-x^2}}{\pi}.$$

Следовательно, $p_1(x)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-x^2}, & |x|\le 1\\ 0, & |x|>1\end{cases}.$

Аналогично: $p_2(y)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-y^2}, & |y|\le 1\\ 0, & |y|>1\end{cases}.$

Для того чтобы установить, являются ли случайные величины X и Y зависимыми, находим условные плотности вероятности:

$$p_y(x)=\begin{cases}\dfrac{1}{2\sqrt{1-y^2}}, & |y|<1, |x|<\sqrt{1-y^2}\\ 0, & |y|<1, |x|>\sqrt{1-y^2}\end{cases}.$$

$$p_x(y)=\begin{cases}\dfrac{1}{2\sqrt{1-x^2}}, & |x|<1,\ |y|<\sqrt{1-x^2}\\ 0, & |x|<1,\ |y|>\sqrt{1-x^2}\end{cases}.$$

Так как $p_1(x)\ne p_y(x)$, $p_2(y)\ne p_x(y)$, то случайные величины X и Y являются зависимыми. Являются ли эти величины коррелированными? Для этого вычислим ковариацию. Имея в виду, что по соображениям симметрии $EX=EY=0$, получим:

$$cov(X,Y)=\iint\limits_{(D)}xyp(xy)dxdy=\frac{1}{\pi}\iint\limits_{(D)}xydxdy=\frac{1}{\pi}\int_{-1}^{1}xdx\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}ydy=$$

$$=\frac{1}{\pi}\int_{-1}^{1} x\frac{y^2}{2}\Big/_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\,dx=\frac{1}{2\pi}\int_{-1}^{1} x\Big[\big(1-x^2\big)-\big(1-x^2\big)\Big]dx=0,$$ т. е. случайные величины X и Y некоррелированы.

Задачи

9.1. Имеется двумерная случайная величина (X, Y) с плотностью $p(x, y)=Axy$ в области D и $p(x, y)=0$ вне этой области. Область D — треугольник, ограниченный прямыми $x+y-1=0$, $x=0$, $y=0$. Найти: а) величину A; б) EX и EY; в) Dx и Dy; г) $cov(x, y)$; д) $r(x, y)$.

Ответ: а) $A=24$; б) $EX=EY=2/5$; в) $Dx=Dy=1/25$;
г) $cov(x, y)=-2/75$; д) $r(x, y)=-2/3$.

9.2. Система случайных величин (X, Y) равномерно распределена в треугольнике, ограниченном прямыми $x+y=a$, $x=a$, $y=a$, где $a>0$. Требуется определить: а) функцию распределения и плотность системы случайных величин (X, Y); б) законы распределения одномерных случайных величин X и Y; в) условные функции и плотности распределения одномерных случайных величин X и Y.

Ответ:

а) $$F(x,y)=\begin{cases}0,\ x\le 0,\ y\le 0,\ x+y\le a\\ \dfrac{1}{a^2}(x+y-a)^2,\ x+y>a,\ x\le a,\ y\le a\\ \dfrac{y^2}{a^2},\ x>a,\ 0<y\le a\\ \dfrac{x^2}{a^2},\ 0<x\le a,\ y>a\\ 1,\ x>a,\ y>a\end{cases}$$

$$p(x,y)=\begin{cases}\dfrac{2}{a^2},\ x+y\ge a,\ x\le a,\ y\le a,\\ 0,\ x+y<a,\ x>a,\ y>a.\end{cases}$$

б) $F_1(x)=\begin{cases}0, \ x\le 0\\ \dfrac{x^2}{a^2}, \ 0<x\le a,\\ 1, \ x>a\end{cases}$ $F_2(y)=\begin{cases}0, \ y\le 0\\ \dfrac{y^2}{a^2}, \ 0<y\le a\\ 1, \ y>a\end{cases}$

$$p_1(x)=\begin{cases}\dfrac{2x}{a^2}, \ 0<x\le a,\\ 0, \ x\le 0, \ x>a\end{cases}, \quad p_2(y)=\begin{cases}\dfrac{2y}{a^2}, \ 0<y\le a,\\ 0, \ y\le 0, \ y>a.\end{cases}$$

в) $p_y(x)=\begin{cases}\dfrac{1}{x}, \ 0<x\le a, \ a-x<y<a\\ 0, \text{ в остальных случаях,}\end{cases}$

$$p_x(y)=\begin{cases}\dfrac{1}{y}, \ 0<y\le a, \ a-y<x<a\\ 0, \text{ в остальных случаях.}\end{cases}$$

9.3. Плотность вероятности системы случайных величин (X, Y) задана выражением: $p(x,y)=\begin{cases}x+y, \ 0<x<1, \ 0<y<1\\ 0, \text{ в остальных случаях.}\end{cases}$

Найти коэффициент корреляции величин X и Y.

Ответ: $-\dfrac{1}{11}$.

9.4. Система случайных величин (X, Y) подчинена закону равномерной плотности распределения внутри квадрата со стороной a, диагонали которого совпадают с осями координат. Установить, являются ли случайные величины X и Y зависимыми. В случае их зависимости установить, являются ли они коррелированными.

Ответ: X и Y зависимые, но некоррелированные случайные величины.

10. Функции случайных величин. Предельные теоремы

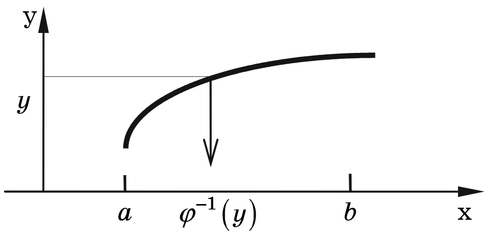
10.1. Функции случайных величин

Если определена функция $\varphi : R_1 \to R_1$, в виде правила φ соответствия элементу (числу) $x \in R_1$ единственного числа $y \in R_1 : y = \varphi(x)$, то можно рассматривать функционально связанными и две случайные величины $Y = \varphi(X)$, где X — случайная величина — аргумент с известной функцией (законом) распределения. Необходимо найти функцию (закон) распределения Y.

Случаи:

1) X — дискретная случайная величина:
$X_1 X_2 \ldots X_n \Rightarrow Y$ — дискретная случайная величина:
$\varphi(X_1) \ldots \varphi(X_n)$
$P_1 P_2 \geq P_n$ соответствие сохраняется $P_1 P_2 \geq P_n$.

2) X — непрерывная случайная величина. Тогда $Y = \varphi(X)$ — непрерывная случайная величина.
   а) $\varphi(x)$ — непрерывна и строго возрастает на $[a, b]$, тогда обратная функция $x = \varphi^{-1}(y)$ — возрастающая.



Ищется $F_Y(y) = P\{Y < y\}$.

$F_Y(y) = P\{Y < y\} = P\left\{X < \varphi^{-1}(y)\right\} = F_X\left(\varphi^{-1}(y)\right)$, т. е. имеем

$$\boxed{F_Y(y) = F_X\left(\varphi^{-1}(y)\right).}$$

Дифференцируя по y обе части равенства, для плотности получим $p_Y(y) = p_X\left(\varphi^{-1}(y)\right)\cdot\left(\varphi^{-1}(y)\right)'$.

б) $\varphi(x)$ — непрерывна, строго убывает. Тогда и $\varphi^{-1}(y)$ убывает, и $F_Y(y) = P\{Y < y\} = P\left\{X > \varphi^{-1}(y)\right\} = 1 - F_X\left(\varphi^{-1}(y)\right)$.

Для плотности имеем $p_Y(y) = -p_X\left(\varphi^{-1}(y)\right)\cdot\left(\varphi^{-1}(y)\right)'$. Но для убывающей функции $\varphi^{-1}(y)$ производная отрицательная. Объединяя случаи а) и б) для монотонной φ имеем:

$p_Y(y) = p_X\left(\varphi^{-1}(y)\right)\cdot\left|\left(\varphi^{-1}(y)\right)'\right|$, т. е. доказана теорема 1.

Т е о р е м а 1. Плотность распределения монотонной функции случайной величины равна произведению плотности случайной величины от обратной функции на модуль производной обратной функции, т. е.

$$\left.\begin{array}{l} X \text{ — непрерывная случайная величина} \\ \varphi(\cdot) \text{ — функциональная связь } Y = \varphi(X) \\ \exists F(x) \text{ — функция распределения для } X \\ \varphi \text{ — монотонная непрерывная функция} \end{array}\right\} \Rightarrow$$

$$\Rightarrow p_Y(y) = p_X\left(\varphi^{-1}(y)\right)\cdot\left|\left(\varphi^{-1}(y)\right)'\right|.$$

З а м е ч а н и е 1. Для кусочно-монотонной функции φ область задания «разбивают» на участки, где φ — монотонна, и тогда имеем

$$p_Y(y) = p_X\left(\varphi_1^{-1}(y)\right)\cdot\left|\left(\varphi_1^{-1}(y)\right)'\right| + p_X\left(\varphi_2^{-1}(y)\right)\cdot\left|\left(\varphi_2^{-1}(y)\right)'\right| + \ldots$$

Пример 1.

$$X \in N\left(a, \sigma^2\right);\ y = \varphi(x) = Ax + B;\ Y = AX + B;\ x = \varphi^{-1}(y) = \frac{y - B}{A}.$$

$$\left|\left(\varphi^{-1}(y)\right)'\right| = \frac{1}{|A|} \therefore p_Y(y) = \frac{1}{|A|\sqrt{2\pi}\cdot\sigma}\cdot e^{-\frac{\left(\frac{y^2-B}{A}-a\right)^2}{2\sigma^2}} = \frac{1}{|A|\sigma\sqrt{2\pi}}\cdot e^{-\frac{(y-B-aA)^2}{2\sigma^2 A^2}}$$

$Y \in N\left(aA+B, \sigma^2|A|\right)^2$. Линейная функция от нормальной случайной величины есть нормальная случайная величина.

Пример 2.

$$X \in N\left(a,\sigma^2\right),\ y = e^x,\ Y = e^X;\ \varphi^{-1}(y) = \ln y;\ \left(\varphi^{-1}(y)\right)' = \frac{1}{y};$$

$$p_Y(y) = \begin{cases} \dfrac{1}{y\sigma\sqrt{2\pi}}\cdot e^{-\frac{(\ln y - a)}{2\sigma^2}}, & y > 0 \\ 0, & y < 0. \end{cases}$$

Пример 3.

$$X \in N(0,1);\ y = x^2;\ Y = X^2;\ \varphi^{-1}(y) = x = \sqrt{y};\ \left(\varphi^{-1}(y)\right)' = \frac{1}{2\sqrt{y}}$$

$$p_Y(y) = \begin{cases} 0, y < 0 \\ \dfrac{2\cdot 1}{2\sqrt{2\pi y}} e^{-\frac{y}{2}} \end{cases}$$ — распределение Пирсона с одной степенью свободы.

10.2. Распределение суммы двух случайных величин

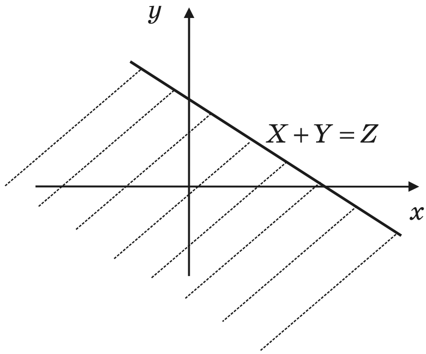
Теорема 2. Функция (плотность) распределения суммы двух независимых непрерывных случайных величин является композицией — сверткой функций (плотностей) распределений слагаемых, т. е.

X и Y непрерывные, независимые случайные величины.
$\exists p(x,y)$; $p_1(x)$; $p_2(y)$
$F(x,y)$; $F_1(x)$; $F_2(y)$
$Z = X + Y$

$$\Rightarrow \quad F(z) = F_1 * F_2(z) = \int_{-\infty}^{\infty} F_1(z-y)\,dF_2(y) = \int_{-\infty}^{\infty} F_2(z-x)\,dF_1(x)$$

$$p(z) = p_1 * p_2(z) = \int_{-\infty}^{\infty} p_2(z-x)\,p_1(x)\,dx = \int_{-\infty}^{\infty} p_1(z-y)\,p_2(y)\,dy$$

(*Композиция — свертка распределений*)



Доказательство.

$p(x,y) = p_1(x)\cdot p_2(y)$. $F(z) = P\{Z < z\} = P\{X + Y < z\}$.

$$F(z) = \int_{-\infty}^{\infty} p_1(x)\,dx \int_{-\infty}^{z-x} p_2(y)\,dy = \int_{-\infty}^{\infty} F_2(z-x)\,p_1(x)\,dx =$$

$$= \int_{-\infty}^{\infty} F_1(z-y)\,p_2(y)\,dy \left(\frac{d}{dz}\right).$$

$$p(z) = \int_{-\infty}^{\infty} p_2(z-x)\,p_1(x)\,dx = \int_{-\infty}^{\infty} p_1(z-y)\,p_2(y)\,dy.$$

Что и требовалось доказать.

Замечание 2. Если хотя бы одна из X и Y непрерывна $\Rightarrow$ $X+Y$ — непрерывна.

Замечание 3. $X \in N\left(a_1, \sigma_1^2\right)$; $Y \in N\left(a_2, \sigma_2^2\right)$ и X, Y независимы $\Rightarrow X+Y \in N\left(a_1+a_2, \sqrt{\sigma_1^2+\sigma_2^2}\right)$.

Замечание 4. Устойчивостью к сложению обладают также распределения Пуассона, Коши, гиперболическое и другие.

10.3. Закон больших чисел. Предельные теоремы

Пусть рассматривается последовательность случайных величин. Как для всякой числовой последовательности имеет смысл рассматривать ее предел. Это значит, что можно ввести понятие сходимости для последовательности случайных величин. Обычно рассматривают следующие виды сходимости.

Определение 1. $\{X_n\}$ — последовательность случайных величин сходится к случайной величине X *по вероятности*, если вероятность уклонения от этой случайной величины стремится к нулю, т. е. $\lim\limits_{n\to\infty} P\left\{\left|X-X_n\right|>\varepsilon\right\}=0$.

Определение 2. Последовательность случайных величин $\{X_n\}$ сходится к случайной величине X в *среднеквадратичном*, если математическое ожидание квадрата уклонения от случайной величины стремится к нулю, т. е. $\lim\limits_{n\to\infty} E\{X_n-X\}^2=0$.

Определение 3. $\{X_n\}$ — последовательность случайных величин сходится к случайной величине X *почти наверное* (с вероятностью 1), если вероятность наибольшего уклонения от случайной величины стремится к нулю, т. е. $\lim\limits_{n\to\infty} P\left\{\sup\limits_{n>n_0}\left|X-X_n\right|>\varepsilon\right\}=0$.

Иногда удобно рассматривать центрированную последовательность случайных величин

$$Y_n = \frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{EX_1 + \ldots + EX_n}{n},\ EX_i < \infty.$$

1. **Закон больших чисел** (теорема Бернулли). Вероятность уклонения центрированной случайной величины от нуля в пределе стремится к нулю, т. е. $Y_n \xrightarrow[n\to\infty]{p} 0$ (по вероятности), или

   $P\{|Y_n| > \varepsilon\} \underset{n\to\infty}{\to} 0.$

2. **Усиленный закон больших чисел** (теорема Бореля). Он соответствует сходимости *почти наверное* последовательности центрированных случайных величин к нулю, т. е.

$$Y_n = \frac{1}{n}\sum_{1}^{n}\left(X_k - EX_k\right) \xrightarrow{\text{п. н.}} 0, \text{ или}$$

$$\left.\begin{array}{l} X_i - \text{независ., одинаково} \\ \text{распределенные} \\ DX_i < \infty \end{array}\right\} \Rightarrow \lim_{n\to\infty} P\left\{ sup\left|\frac{\mu}{n} - p\right| > \varepsilon \right\} = 0, \text{ где}$$

   $\mu = \sum X_i.$

3. **Неравенство Чебышева.** Вероятность уклонения случайной величины с конечной дисперсией от математического ожидания ограничена отношением дисперсии к квадрату уклонения, т. е. X — случайная величина; EX, $DX < \infty \Rightarrow$

$$\boxed{\Rightarrow P\{|X - EX| > \varepsilon\} \le \frac{DX}{\varepsilon^2}.}$$

4. **Теорема Чебышева.** Для последовательности случайных величин $\{X_n\}$ применим закон больших чисел, если случайные величины X_k — попарно независимы и равномерно ограничены (одним и тем же числом), т. е. $DX_k < L$.

5. **Теорема Маркова.** Для того чтобы к последовательности случайных величин $\{X_n\}$ был применим закон больших чисел, достаточно, чтобы $\lim \frac{1}{n^2} D\left(\sum_1^n X_k\right) = 0$, т. е.

$$\left.\begin{array}{c} X_k \text{ — произвольная с.в., } EX_k < \infty;\ DX_k < \infty \\ \lim\limits_{n\to\infty} \dfrac{D\sum\limits_1^n X_k}{n^2} = 0 \end{array}\right\} \Rightarrow$$ к $\{X_n\}$ применим закон больших чисел, $P\{|Y_n| > \varepsilon\} \underset{n\to\infty}{\to} 0.$

6. **Теорема.** Чтобы к последовательности произвольных случайных величин $\{X_n\}$ был применим закон больших чисел, необходимо и достаточно, чтобы ($\Leftrightarrow$) имело место:

$$\lim_{n\to\infty} E \frac{\left(\sum\limits_1^n (X_k - EX_k)\right)^2}{n^2 + \left(\sum\limits_1^n (X_k - EX_k)\right)^2} = 0.$$

7. **Теорема Ляпунова.** Для того чтобы уклонение суммы независимых случайных величин было в пределе распределено *нормально*, достаточно, чтобы выполнялось условие

$$\lim_{n\to\infty} \frac{\sum\limits_1^n C_k}{B_k^{2+\sigma}} = 0, \text{ где } \begin{array}{c} C_k \equiv E|X_k - EX_k|^{2+\sigma}, \sigma > 0 \\ B_n^2 \equiv \sum\limits_1^n DX_k \end{array}.$$

Тогда

$$\lim_{n\to\infty} P\left\{a < \frac{1}{B_n}\sum_1^n (X_k - EX_k) < b\right\} = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{t^2}{2}} dt = \left[\Phi(b) - \Phi(a)\right].$$

10.4. Примеры и задачи

Примеры

Пример 10.1. Дан ряд распределения случайной величины X.

X_i	0	1	2
P_i	0,2	0,3	0,5

Найти распределение случайной величины
$Y=\varphi(X)=X^3-2X^2+X-1$.

Решение: Определяем возможные значения Y и их вероятности:
$y=\varphi(0)=-1$; $y=\varphi(1)=-1$; $y=\varphi(2)=1 \Rightarrow$
$y_1=-1$; $P(Y=-1)=P(X=0)+P(X=1)=0,2+0,3=0,5$,
$y_2=1$; $P(Y=1)=P(X=2)=0,5$.

Имеем распределение случайной величины Y:

Y_i	–1	1
P_i	0,5	0,5

Пример 10.2. Найти плотность распределения случайной величины $Y=5-2X$, если случайная величина X имеет нормальное распределение с $EX=0$, $\sigma=4$, т. е. $X\in N(0, 4)$.

Решение: Обратная функция $\varphi^{-1}(y)=(5-y)\cdot\dfrac{1}{2}$.

$$\left|\left(\varphi^{-1}(y)\right)'\right|=\frac{1}{2}$$

$$p_y(y)=\frac{1}{2}p_x\left(\frac{5-y}{2}\right)=\frac{1}{8\sqrt{2\pi}}e^{-\frac{(5-y)^2}{128}}=\frac{1}{8\sqrt{2\pi}}e^{-\frac{(y-5)^2}{128}}.$$

Задачи

10.1. Найти плотность распределения случайной величины $Y=\sqrt{X}$, если X — случайная величина с плотностью распределения $p_x(x)$ и возможными значениями от 0 до ∞.

Ответ: $p_y(y)=2y\cdot p_x(y^2)$, $y>0$.

10.2. Случайная величина X распределена равномерно в интервале $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$. Найти плотность распределения случайной величины $Y=\sin X$.

Ответ: $p_y(y)=\dfrac{1}{\pi\sqrt{1-y^2}}$, $y\in(-1{,}1)$; $p_Y(y)=0$ для $y\notin(-1{,}1)$.

10.3. Найти плотность распределения суммы двух независимых случайных величин X и Y, имеющих нормальное распределение с математическими ожиданиями 0 и 2, дисперсиями 6 и 3, т. е. $X\in N\left(0,\sqrt{6}\right)$, $Y\in N\left(2,\sqrt{3}\right)$.

Ответ: $Z=X+Y$, $p(z)=\dfrac{1}{3\sqrt{2\pi}}e^{-\frac{(z-2)^2}{18}}$.

10.4. Составить композицию двух показательных законов с параметрами $\lambda_1=\lambda_2=\lambda=0{,}2$ (т. е. найти закон распределения суммы двух независимых случайных величин X_1 и X_2, имеющих плотности: $p_1(x_1)=\lambda e^{-\lambda x_1}$, $p_2(x_2)=\lambda e^{-\lambda x_{21}}$ $(x_2>0,\ x_1>0)$.

Ответ: $Z=X_1+X_2$, $p(z)=\lambda^2 e^{-\lambda z}\big|_{\lambda=0,2}=0{,}004e^{-0,2z}$.

10.5. Дискретная случайная величина X имеет ряд распределения:

X_i	2	–1	0	1	2
P_i	0,1	0,2	0,3	0,3	0,1

Построить ряд распределений случайной величины $Y = X^2 + 1$.
Ответ:

Y_i	1	2	5
P_i	0,3	0,5	0,2

10.6. Случайная величина X распределена по нормальному закону с характеристиками EX и σ_x^2. Случайные величины Y и Z связаны с X зависимостями $Y = X^2$, $Z = X^3$. Найти ковариации (корреляционные моменты): $cov(X,Y) = K_{xy}$, $cov(X,Z) = K_{xz}$, $cov(Y,Z) = K_{yz}$.

Ответ: $K_{xy} = 2\sigma^2 m_x$; $K_{xz} = 3\sigma_x^4 + 3m_x^2\sigma_x^2$; $K_{yz} = 12m_x\sigma_x^4 + 6m_x^3\sigma_x^2$.

11. Основные понятия из теории случайных процессов

Случайные величины часто оказываются довольно грубыми моделями для описания наблюдений, особенно при наблюдениях за поведением динамических систем. Если требуется иметь более строгие и тонкие обоснования с количественными оценками, то необходимо использовать другие объекты теории вероятностей. Наиболее адекватной математической моделью при описании явлений в динамике являются случайные процессы.

11.1. Определение случайного процесса

Определение 1. Под *случайным процессом* $X_t \equiv X(t)$ понимают однопараметрическое семейство случайных величин, где параметр t принадлежит произвольному множеству вещественных чисел, т. е. $t \in \mathrm{T} \subset R_1$. Этот параметр называют временем. Если параметр t берется из счетного множества, то случайный процесс называют *дискретным*, а если из континуального, то *непрерывным*.

Можно считать, что случайный процесс есть индексированная случайная величина. На языке математики случайный процесс есть числовая функция, заданная на произведении множеств (σ-алгебры и множества T), т. е. $X_t \equiv f(\omega, t)$ есть функция двух переменных. При фиксированном параметре $t \in \mathrm{T}$ имеет место случайная величина в точке t, а при фиксированном событии $\omega_0 \in \Omega$ имеет место неслучайная функция $f(\omega_0, t)$, которую называют *выборочной функцией* или *реализацией*.

1. $T=\{\,.\,\}$, тогда случайный процесс X есть случайная величина.
2. $T=\{1,\ 2,\ \ldots,\ n\}$, тогда случайный процесс есть случайный вектор размера n: $(X_1,\ X_2,\ \geq,\ X_n)$.
3. $T=\{1,\ 2,\ \ldots,\ n,\ \ldots\}$, тогда случайный процесс есть последовательность случайных величин: $(X_1,\ X_2,\ \geq,\ X_n,\ \geq)$.
4. Имеется n случайных величин $X_1,\ X_2,\ \geq,\ X_n$ и n вещественных чисел $\lambda_1,\ \lambda_2, \ldots, \lambda_n$, тогда $X(t)=\sum\limits_{i=1}^{n} \mathrm{X}_i \cos\lambda_i t$ есть случайный процесс («гармоники» со случайными амплитудами).

11.2. Простейшие характеристики случайного процесса

Случайный процесс обычно описывается через семейство конечномерных распределений, удовлетворяющих специальным условиям (симметрии и согласования). Случайный процесс можно рассматривать как обобщение n-мерного случайного вектора. Покажем это. Пусть имеется вероятностное пространство $\{\Omega, \Xi, P\}$, на котором задан случайный вектор $(X_1,\ X_2,\ \geq,\ X_n)$. *Функцией распределения (распределением)* случайного вектора называют $F(x_1, x_2, \ldots, x_n) = P\{(X_1, X_2, \ldots, X_n) \in A\}$, где A — есть борелево множество (из σ-алгебры Ξ). Из числовых характеристик наиболее информативными являются моменты:

1) Начальный момент 1-го порядка (математическое ожидание) — $EX = (EX_1,\ EX_2,\ \geq,\ EX_n)^T$;
2) Смешанный момент 2-го порядка (начальный и центральный) — *ковариационная* матрица $R_1 = \{E(\mathrm{X}_i \mathrm{X}_j)\}$, $i,\ j=1, 2,\ \ldots,\ n$ или *корреляционная* матрица $R=\{E(X_i - EX_i)(X_j - EX_j)\}$, i, $j=1,\ 2,\ \ldots,\ n$.

Для случайного процесса имеем по аналогии следующее. На том же вероятностном пространстве задается случайный процесс $X(t) = X(\omega, t)$, $t \in T$.

Далее из набора всевозможных конечномерных распределений, соответствующих произвольным случайным векторам

$\left(X(t_1), X(t_2), \dots, X(t_n)\right)$ вида

$$\left\{\begin{array}{c} F\left(x_{t_1}, x_{t_2}, \dots, x_{t_n}\right) = P\left\{\left(X(t_1), X(t_2), \dots, X(t_n)\right) \in A\right\} \\ t_1,\ t_2, \dots,\ t_n \in T;\ n = 1, 2, 3, \dots \end{array}\right\},$$ выбирается распределение, удовлетворяющее специальным условиям симметрии и согласования.

Соответствующие числовые характеристики для случайного процесса:

1) Математическое ожидание — среднее, если оно существует $a(t) = EX(t)$.
2) *Корреляционная функция* (смешанный центральный момент 2-го порядка): $R(t, \tau) = E\left(X(t) - a(t)\right)\left(X(\tau) - a(\tau)\right)$, или *ковариационная функция* (смешанный начальный момент 2-го порядка): $R_1(t, \tau) = EX(t)X(\tau)$.

На практике эти характеристики используются очень широко.

11.3. О некоторых типах случайных процессов

Существует довольно большое количество разных типов случайных процессов. В математической литературе имеются их классификации по различным признакам. Наиболее простыми и распространенными являются следующие типы.

Определение 2. Случайный процесс $X(t)$ называется *стационарным в широком смысле,* если его математическое ожидание есть постоянная величина, т. е. $a(t) = const$, а корреляционная функция ограничена и зависит только от разности аргументов: $R(t, \tau) = R(t - \tau)$.

Определение 3. Стационарный случайный процесс $X(t)$ называется *эргодическим (с сильным перемешиванием)*, если его среднее значение по времени (на интервале T) равно математическому ожиданию.

Определение 4. Случайный процесс $X(t)$ называется *гауссовским*, если все его конечномерные распределения гауссовские, т. е.

$$p(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{detR}} exp\left\{-\frac{1}{2}\left(R^{-1}x, x\right)\right\}, \text{ где } R = \left\{r_{ij}\right\}_{n\times n},$$

$r_{ij} = E(X_i - a_i)(X_j - a_j)$, $a_i = EX_i$, $r_{ii} = DX_i$, $detR > 0$, причем для любых $t_1, t_2, \ldots, t_n \in T$.

Для гауссовского случайного процесса имеет место $a(t) = EX(t)$, $R(t,\tau) = E\left(X(t) - a(t)\right)\left(X(\tau) - a(\tau)\right)$.

Если заданы $a(t)$ и $R(t,\tau)$, то для любого случайного вектора $\left(X(t_1), X(t_2), \ldots, X(t_n)\right)$ задано математическое ожидание (среднее значение) и корреляционная матрица R размера $n\times n$. Следовательно, $a(t)$ и $R(t,\tau)$ полностью определяют все конечномерные распределения, а потому и полностью характеризуют гауссовский случайный процесс.

Определение 5. Случайный процесс $X(t)$ называется *с независимыми приращениями,* если для любых t_i, $i = 1, 2, \geq, n$ таких, что $t_1 \le t_2 \le \ldots \le t_i \le t_{i+1} \le \ldots \le t_n$, случайные величины-приращения $\left\{\left(X(t_2) - X(t_1)\right), \left(X(t_3) - X(t_2)\right), \ldots, \left(X(t_n) - X(t_{n-1})\right)\right\}$ независимы, а если, кроме того, приращение $\left(X(t_i + h) - X(t_i)\right)$ зависит только от разности h и не зависит от величины t_i, то такой процесс называют *с независимыми и однородными приращениями*.

Определение 6. Процессом *броуновского движения (винеровским)* называют гауссов процесс с независимыми и однородными приращениями, т. е. такой, что для любых $t > s$ приращение $X(t) - X(s)$ распределено по нормальному закону с дисперсией $\sigma^2 \cdot (t - s)$.

На практике часто используется модель в виде случайного процесса типа броуновского движения со смещением. Такой процесс имеет специальное название и обычно является адекватной моделью при описании динамических систем со случайными возмущениями.

Определение 7. Пусть $X(t)$ есть случайный процесс, элементарное приращение которого представляется в виде двух слагаемых, т. е. $X(t+dt)-X(t)=a(t,X(t))\cdot dt+b(t,X(t))(W(t+dt)-W(t))$, где $W(t)$ — процесс броуновского движения, $a(t,X(t))$ — коэффициент смещения (сноса), $b(t,X(t))$ — коэффициент диффузии. Тогда случайный процесс $X(t)$, удовлетворяющий условию $dX(t)=$ $=a(t,X)dt+b(t,X)dW$, называется *диффузионным* процессом (процессом «белого шума»).

Довольно широкое распространение в построении математических моделей динамических систем со случайными возмущениями получили также *марковские* случайные процессы, которые можно рассматривать, по существу, как обобщение марковских цепей. С другой стороны, частным случаем марковских процессов являются и диффузионные процессы. Классу марковских процессов посвящена обширная литература.

Часть II. Математическая статистика

12. Статистические распределения

12.1. Предмет математической статистики и статистические совокупности

Выявление и исследование закономерностей, которым подчиняются реальные процессы, является сутью любой научной дисциплины. В социально-экономических дисциплинах закономерности, как правило, выявляются с помощью *целенаправленного статистического изучения* массовых явлений, включающего следующие этапы:

1) сбор данных;
2) систематизацию и упорядочение;
3) статистический анализ.

Предмет математической статистики составляют приемы и способы научного анализа данных, относящихся к массовым явлениям, с целью определения обобщающих характеристик и выявления *статистических закономерностей*, обладающих статистической устойчивостью.

Связь вероятности и статистики. Изучение вероятностных моделей позволяет понять различные свойства случайных событий на *абстрактном* уровне, не прибегая к практике (эксперименту). В математической статистике исследование связано с конкретными данными и идет от практики (наблюдения) к гипотезе и ее проверке.

Теорию случайных величин следует рассматривать как систему математических предложений, которая может служить моделью явления статистической устойчивости, наблюдаемого в связи с последовательностями случайных экспериментов.

Определение 1. Множество *однородных* объектов $\{\Re_i\}_1^n$, подлежащих статистическому изучению на основе случайного эксперимента, эквивалентного схеме равновероятного выбора элементов из множества с возвращением, называется *статистической совокупностью*; отдельные объекты $\Re_i$ — *элементами* совокупности, а их число (n) — *объемом* совокупности. Статистическую совокупность также можно рассматривать как множество реализаций в схеме независимых испытаний для некоторой случайной величины, которую и принимают за математическую модель.

Примеры: 1. $\{\Re_i\}$ — рабочие предприятия; 2. $\{\Re_i\}$ — предприятия региона.

Определение 2. Под *признаком* (A) понимают ту или иную характеристику элементов совокупности, описывающую некоторое свойство или состояние наблюдаемого элемента. Признак может быть *количественным* и *качественным*.

Примеры: 1. Пол, профессия, сорт, цвет и т. д. — *качественные* признаки. 2. Масса, рост, объем, заработная плата, прибыль и т. д. — *количественные* признаки.

Замечание 1. Особая группа признаков — *ранговые* показатели. Их значения устанавливаются путем упорядоченного ранжирования объектов совокупности в соответствии с некоторым признаком и присвоения каждому из объектов порядкового номера **ранга**.

Замечание 2. Количественные признаки могут быть *дискретными* и *непрерывными* (например, результаты счета (g) и результаты измерений (n)).

Определение 3. Процесс изучения признаков элементов совокупности (измерение, регистрация, описание и т. д.) называют ***статистическим наблюдением***. *Сплошное наблюдение* — изучается каждый элемент совокупности, *выборочное наблюдение* — изучается выборка, т. е. часть совокупности (*выборочная совокупность*). В последнем случае всю исходную совокупность называют ***генеральной***.

Расчленение совокупности, при котором группы будут состоять из однородных элементов, называют ***группировкой***.

Порядок организации статистического наблюдения:

а) формулировка задачи;
б) определение признака, который изучается;
в) описание границ статистической совокупности;
г) задание способа проведения наблюдений (сплошное или выборочное);
д) выбор правила отбора элементов совокупности;
е) выбор способа регистрации данных и задание необходимой точности измерений.

Основные требования к совокупности:

1) однородность;
2) сопоставимость, т. е. практическая неразличимость элементов по другим, кроме изучаемого, признакам.

12.2. Распределение качественных признаков

Определение 4. Пусть $\{\Re_i\}_1^n$ — совокупность, A — альтернативный признак, т. е. A — наличие признака, $\bar{A}$ — отсутствие признака для изучаемых элементов $\Re_i$. Тогда результаты наблюдения, представляемые в виде 2-х разрядной группировки

$\Re_i$	1	2	3	4	… n
Результат наблюдений	A	$\bar{A}$	A	$\bar{A}$	… $\bar{A}$

Варианта	A	$\bar{A}$	Σ
Частота	m_A	$m_{\bar{A}}$	n

$: m_A + m_{\bar{A}} = n,$

называют *распределением альтернативного признака*.

Определение 5. Относительные частоты $\frac{m_A}{n}$ и $\frac{m_{\bar{A}}}{n}$ называют *долями признака*.

Замечание 3. Если рассматривается несколько вариаций признака A, то результаты наблюдений представляют в виде *многоразрядной* группировки

Варианта	A_1	A_2	$\geq A_S$	Σ
Частота	m_1	m_2	$\geq m_S$	n

, где $\sum_1^S m_i = n,$

которую также называют *распределением признака А*.

Определение 6. Пусть $\{\Re_i\}_1^n$ — совокупность; A и B — альтернативные признаки для элементов. Тогда результаты наблюдения в виде *группировки 2-го порядка* называют распределением двух альтернативных признаков:

B \ A	B	$\bar{B}$	Σ
A	m_{AB}	$m_{A\bar{B}}$	m_A
$\bar{A}$	$m_{\bar{A}B}$	$m_{\bar{A}\bar{B}}$	$m_{\bar{A}}$
Σ	m_B	$m_{\bar{B}}$	n

$$\begin{cases} m_A + m_{\bar{A}} = n; \qquad m_B + m_{\bar{B}} = n \\ m_{AB} + m_{A\bar{B}} = m_A;\ m_{AB} + m_{\bar{A}B} = m_B \\ m_{\bar{A}B} + m_{\bar{A}\bar{B}} = m_{\bar{A}};\ m_{A\bar{B}} + m_{\bar{A}\bar{B}} = m_{\bar{B}} \end{cases}$$

Замечание 4. Если признаки A и B имеют вариации по группам s и t, то группировка 2-го порядка для наблюдений представляет собой матрицу $s \times t$ порядка, которую называют *таблицей взаимосопряженности*.

12.3. Распределение количественных признаков

Определение 7. Каждое отдельное значение признака X элементов называют ***вариантой*** X_i, а изменчивость величины признака ***вариацией***.

Определение 8. Группировка совокупности по отдельным вариантам называется *первичной обработкой* данных наблюдений. При этом число повторений m_i вариант X_i называют *частотой* варианты X_i (статистический вес), а относительную частоту $w_i = m_i/n$ — *частостью*.

Определение 9. Упорядоченная совокупность вариант признака ($X_1 < X_2 < \geq < X_n$) с учетом их частоты называется дискретным *вариационным рядом*, который представляют обычно двумя строками: 1-я — варианты, 2-я — частоты.

Имеют место: $\sum_{i=1}^{n} m_i = n$ и $\sum_{i=1}^{n} w_i = 1$. Множество точек (X, m_i) образуют *точечную диаграмму распределения*.

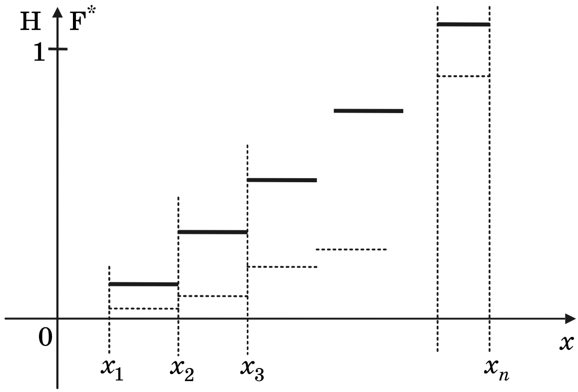
X	X_1	X_2	X_3	$\geq$	X_n	Σ
m	m_1	m_2	m_3	$\geq$	m_n	n

Определение 10. Сумма частоты данной варианты с суммой частот всех предшествующих вариант признака X называется *накопленной (кумулятивной) частотой* $H(x)$. Другими словами, это число элементов совокупности, имеющих значение при-

знака X меньше заданного числа. Аналогично определяется *накопленная (кумулятивная) частость* $F_n^*(x)$. Последнюю обычно называют ***эмпирической функцией распределения***.

$$F_n^*(x) = \begin{cases} 0, \ x < X_1 \\ w_1, X_1 \le x < X_2 \\ w_1 + w_2, X_2 \le x < X_3 \\ \qquad \vdots \\ \sum_{i=1}^{n} w_i, \ x \ge X_n \end{cases}$$

$$F_n^*(x) = \frac{1}{n} \sum_{i}^{n} I_{\{x_i < x\}}$$



З а м е ч а н и е 5. При *большом* числе различных вариант весь диапазон изменения признака разбивают на интервалы и результаты группировки сводят к ***интервальному*** вариационному ряду, в котором частоты относятся не к отдельным вариантам, а ко всему интервалу (X_{i-1}, X_i), т. е. $(X_{i-1} \le x < X_i)$.

Замечание 6. Иногда для анализа интервального вариационного ряда вводят понятие *плотности частоты*: $p_i = \dfrac{m_i}{n\Delta x_i}$.

Графические представления вариационных рядов:

1) **Полигон** распределения есть изображение *дискретного* вариационного ряда в виде ломаной, где отрезки прямых последовательно соединяют точки на плоскости координат: абсцисса — варианта, ордината — частота.
2) **Гистограмма** есть изображение *интервального* вариационного ряда в виде столбцового (ступенчатого) графика — столбограммы, по оси ординат — частоты, по оси абсцисс — интервалы.
3) **Кумулята** (кумулянта) есть изображение интервального вариационного ряда с накопленными частотами (частостями), где ось абсцисс — интервалы варианты, ось ординат — накопленные частоты.
4) **Огива** есть изображение интервального вариационного ряда с накопленными частотами, где ось абсцисс — накопленные частоты, ось ординат — интервалы вариант.

12.4. Числовые характеристики опытных распределений

Как и случайные величины, статистические совокупности имеют числовые характеристики, которые условно делятся на:

1) средние или центральной тенденции;
2) рассеяния;
3) моменты распределения и формы.

(1) Средние характеристики типа

$$\bar{X}_\nu = \sqrt[\nu]{\sum X_i^\nu \cdot m_i \Big/ \sum m_i} = \sqrt[\nu]{\sum X_i^\nu \cdot w_i} \text{ — *среднестепенная*,}$$

где $X_i > 0$.

а) $\nu = 1 \Rightarrow \bar{X}_a = \dfrac{1}{n}\sum_{i=1}^{n} X_i m_i$ — *средняя арифметическая*, где X_i — значение признака в дискретном вариационном ряде или середины интервалов в интервальном вариационном ряде.

б) $\nu = -1 \Rightarrow \bar{X}_{-1} = \dfrac{\sum m_i}{\sum \dfrac{m_i}{X_i}}$ — *средняя гармоническая*

$$\left(= \frac{n}{\sum \frac{1}{X_i}} \text{ при } m_i = 1 \right).$$

в) $\nu = 2 \Rightarrow \bar{X}_2 = \left(\dfrac{\sum X_i^2 \cdot m_i}{\sum m_i} \right)^{\frac{1}{2}}$ — *средняя квадратическая.*

г) $\nu = 0 \Rightarrow \bar{X}_0 = \sqrt[n]{X_1^{m_1} \cdot X_2^{m_2} \cdot \ldots \cdot X_n^{m_n}} = \sqrt[n]{\prod_i X_i^{m_i}}$ — *средняя геометрическая* (при анализе временных рядов).

Свойство *мажорантности* средних: $\boxed{\bar{X}_{-1} \le \bar{X}_0 \le \bar{X}_1 \le \bar{X}_2 \le \ldots}$ (Теорема Боярского).

Замечание 7. По аналогии со случайными величинами применяют **структурные** или порядковые средние: ***медиану*** и ***моду***, пользуясь эмпирической функцией распределения $F^*(x)$ и максимальной частостью соответственно.

M_e — значение признака, приходящегося на середину ранжированного вариационного ряда.

M_o — значение признака, наблюдаемого наибольшее число раз.

(2) Характеристики рассеяния

а) *Размах* варьирования $R = X_{\max} - X_{min}$.

б) *Среднее линейное отклонение* $E_a = \dfrac{1}{n} \sum_{i=1}^{n} \left| X_i - \bar{X} \right| m_i$.

в) *Дисперсия* $D = \sigma^2 = \dfrac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2 m_i$.

г) *Среднее квадратичное отклонение* $\sigma = \sqrt{D}$.

д) *Коэффициент вариации* $V = \frac{\sigma}{\bar{X}} 100\%$ при $\bar{X} \neq 0$.

(3) Моменты распределения, характеристики формы

а) Момент k-го порядка: $\boxed{M_k(c) = \frac{1}{n}\sum_{i=1}^{n}(X_i - c)^k m_i}$.

б) При $c = 0$ имеем *начальный* момент: $\alpha_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k m_i$.

в) При $c = \bar{X}_a$ имеем *центральный* момент:

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_a\right)^k m_i.$$

г) *Асимметрия* $A = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}}$.

д) *Эксцесс* (крутизна) $Э = \frac{\mu_4}{\sigma^4} - 3 = \frac{\mu_4}{\mu_2^2} - 3$.

Замечание 8. Для числовых характеристик **(1)–(3)** справедливы все утверждения для числовых характеристик случайных величин, рассмотренные ранее.

Замечание 9. Если совокупность состоит из отдельных групп, то имеется зависимость между общими и групповыми средними и дисперсиями:

$$\boxed{\bar{X} = \frac{1}{n}\sum_{k=1}^{\ell}\bar{X}_k \cdot n_k} \text{ и } \boxed{D = \frac{1}{n}\sum_{k=1}^{\ell} D_k \cdot n_k + D_{м\,гр}\left(\bar{X}_k\right)}, \text{ где}$$

ℓ — число групп, n_k — численность групп $\left(n = \sum_{k=1}^{\ell} n_k\right)$,

$\bar{X}_k$ и D_k — групповые характеристики,

$D_{м\,гр}\left(\bar{X}_k\right)$ — межгрупповая дисперсия.

$$D_{м\,гр}\left(\bar{X}_k\right) = \frac{1}{n}\sum_{k=1}^{\ell}\left(\bar{X}_k - \bar{X}\right)^2 m_k.$$

13. Введение в теорию выборочного метода

13.1. Выборочные наблюдения

Изучение всей генеральной совокупности, как правило, связано с большими затратами. Поэтому чаще всего изучают только часть совокупности из генеральной, а это связано с решением следующих задач:

1) как организовать выборочное наблюдение, чтобы получить наибольшую информацию (полную) — проблема ***репрезентативности*** выборки;
2) как наилучшим образом использовать результаты выборки проблема ***оценки***.

По результатам выборочного наблюдения требуется высказать определенные суждения о генеральном распределении, по которому обычно делаются некоторые априорные предположения.

Способы формирования выборки:

1) *Случайный* отбор — путем жеребьевки или с помощью таблицы случайных чисел.
2) *Неслучайный* отбор, например, *серийный* (гнездовой) — случайным образом отбираются группы, *механический* — упорядоченные элементы отбираются по установленному признаку. Всегда стремятся к случайному отбору.

Случайный отбор производится по схеме **возвратной** и **безвозвратной** выборки (возвращенный и невозвращенный шар).

13.2. Статистические оценки и требования к ним

Поскольку генеральную совокупность можно рассматривать как множество реализаций (наблюдений) некоторой случайной величины с функцией распределения $F(x)$, то эту функцию и будем называть *функцией распределения* генеральной совокупности. Задача заключается в том, чтобы по результатам одной случайной выборки выполнить следующее:

1) оценить достаточно точно значения неизвестных характеристик (параметров) генерального распределения;
2) либо оценить интервал, в котором находится значение характеристики (параметра) с заданной вероятностью.

В силу случайности отбора возвратную выборку объема n можно рассматривать как схему n повторных независимых испытаний, где результаты каждого испытания $\left(j \in \left[\overline{1,n}\right]\right)$ есть независимые одинаково *с генеральной совокупностью распределенные* случайные величины X_j, т. е. $F_j(x_j, \theta) = F(x, \theta)$ и $F(x_{(1)}, \geq, x_{(n)}, \theta) = F(x_{(1)}, \theta) \ldots F(x_{(n)}, \theta)$.

Определение 1. Пусть $F(x, \theta)$ — функция распределения генеральной совокупности с неизвестным параметром θ. Измеримая функция результатов измерения — выборки $\hat{\theta}_n(X_1, \ldots, X_n)$, значение которой принимается за наилучшее в некотором смысле приближения к значению параметра θ, называется ***точечной оценкой*** параметра θ.

В силу случайности выборки точечная оценка есть случайная величина.

Замечание 1. Измеримую функцию результатов наблюдений $f(X_1, \geq, X_n)$ называют ***статистикой***. Поэтому $\hat{\theta}_n$ — есть статистика.

Примеры точечных оценок:

1. *Выборочная средняя* $\hat{\hat{X}}_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$;

2. *Выборочная дисперсия* $s_n^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}\left(X_i - \hat{\hat{X}}_n\right)^2$.

Определение 2. Точечная оценка $\hat{\theta}_n$ параметра называется ***состоятельной***, если она сходится по вероятности к истинному значению параметра, т. е. $\hat{\theta}_n \xrightarrow{Вер} \theta$, или $\lim\limits_{n\to\infty} P\left\{\left|\hat{\theta}_n - \theta\right| < \varepsilon\right\} = 1$ для $\forall \varepsilon > 0$.

Определение 3. Точечная оценка $\hat{\theta}_n$ — называется ***асимптотически несмещенной***, если предел математических ожиданий оценок совпадает с истинным значением параметра, т. е. $\lim\limits_{n\to\infty} E\hat{\theta}_n = \theta$, и ***несмещенной***, если соответственно $E\hat{\theta}_n = \theta$.

Определение 4. Несмещенную точечную оценку $\hat{\theta}_n$ называют ***эффективной*** (допустимой), если дисперсия оценки является минимальной (из дисперсий всех несмещенных оценок этого параметра), т. е. $D\hat{\theta}_n = \min\limits_{\tilde{\theta}_n} D\tilde{\theta}_n$, где $\tilde{\theta}_n$ — любая несмещенная оценка θ при $F_n^*(x, \theta)$.

Эта оценка реализует равенство в выражении

$$D\hat{\theta}_n \geq \frac{1}{n \cdot E\left(\dfrac{\partial \ln p(x,\theta)}{\partial \theta}\right)^2}$$

— неравенство Рао-Крамера (для одинаково распределенных независимых выборок).

количество информации Фишера

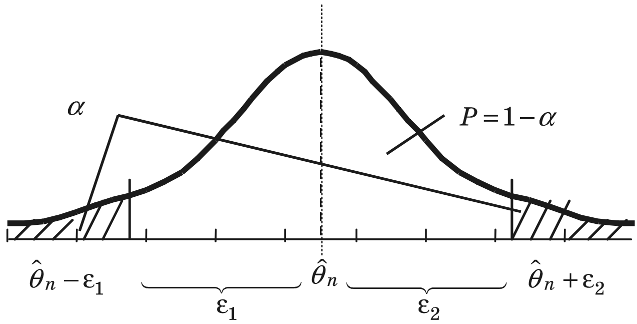
Определение 5. Статистика — точечная оценка $\hat{\theta}_n$ — называется ***достаточной***, если выборочные данные не могут дать дополнительной информации о параметре θ. Иначе: условная плотность вероятности выборки при известном $\hat{\theta}_n$ не зависит от θ.

Требования, предъявляемые к точечным статистическим оценкам:

а) состоятельность;
б) несмещенность;
в) эффективность;
г) достаточность.

Замечание 2. В силу случайного характера выборки точечная оценка $\hat{\theta}_n$ есть случайная величина, а тогда величина $\delta \equiv \theta - \hat{\theta}_n$, называемая *ошибкой выборки*, тоже есть случайная величина.

Определение 6. Пусть $\hat{\theta}_n$ — есть случайная величина, имеющая плотность распределения $p(\theta)$, тогда случайная величина $\delta \equiv \theta - \hat{\theta}_n$ имеет то же распределение. Если задана вероятность $P(-\varepsilon_1 < \delta < \varepsilon_2) = 1 - a$, где $(-\varepsilon_1, \varepsilon_2)$ — интервал, а величина α — близкая к нулю ($\alpha > 0$), то с учетом значения δ эту вероятность можно переписать как $P\left(\hat{\theta}_n - \varepsilon_1 < \theta < \hat{\theta}_n + \varepsilon_2\right) = 1 - \alpha$. Тогда интервал $\left(\hat{\theta}_n - \varepsilon_1,\ \hat{\theta}_n + \varepsilon_2\right)$ называется ***доверительным интервалом*** с доверительными границами $\left(\hat{\theta}_n - \varepsilon_1\right)$ и $\left(\hat{\theta}_n + \varepsilon_2\right)$; вероятность $P = 1 - \alpha$ называется ***доверительной вероятностью***, α — ***уровнем значимости***.



Замечание 3. В отличие от вероятности попадания случайной величины в интервал здесь имеет место вероятность включения неизвестной постоянной величины θ в интервал со случайными границами.

Определение 7. Доверительный интервал $\left(\hat{\theta}_n-\varepsilon_1,\ \hat{\theta}_n+\varepsilon_2\right)$ называют ***интервальной оценкой*** параметра θ.

Таким образом, под интервальной оценкой понимают интервал со случайными границами, который включает в себя неизвестный оцениваемый параметр с заданной вероятностью.

Замечание 4. Если для получения точечной оценки $\hat{\theta}_n$ нужно знать вид функции от выборки, то для интервальной нужно знать еще закон распределения $\hat{\theta}_n$, который обычно задается.

Определение 8. При симметричной функции распределения для $\hat{\theta}_n$ относительно θ можно рассматривать симметричный интервал (ε, ε), т. е. $P\left(\left|\theta-\hat{\theta}_n\right|<\varepsilon\right)=1-\alpha$, тогда ε называют ***предельной ошибкой выборки***.

Замечание 5. При интервальной оценке могут рассматриваться задачи:

1) определение доверительной вероятности по заданному интервалу и объему выборки;
2) определение доверительного интервала по заданной доверительной вероятности и объему выборки;
3) определение объема выборки по заданной доверительной вероятности и доверительному интервалу.

14. Методы построения статистических оценок

14.1. Методы нахождения оценок

Существуют два подхода к выбору функции от наблюдений для оценки параметра генеральной совокупности при удовлетворении требований к статистике:

1) из заданных требований выводится формула для построения оценки (обычно задается вид функции). Например, класс *линейных оценок* $\theta = \sum a_i X_i$. Достаточно предположить существование и конечность генеральной средней и дисперсии;
2) статистика выбирается по некоторым соображениям, а затем проверяются требования.

При последнем подходе используются методы выбора функций:

(1) *Метод аналогии*. Для оценки параметров генеральной совокупности выбираются аналогичные параметры выборочного распределения. Для оценок доли признака, средней и дисперсии генеральной совокупности используется выборочная доля $\hat{\theta} = \hat{p} = \dfrac{m}{n}$, выборочная средняя $\hat{\bar{X}}_n = \dfrac{1}{n}\sum_i X_i$, выборочная дисперсия: $s_n^2 = \dfrac{1}{n}\sum_i^n \left(X_i - \hat{\bar{X}}_n\right)^2$. При проверке требований оценка s_n^2 оказывается смещенной и поэтому должна быть умножена на корректирующий множитель $\dfrac{n}{n-1}$.

(2) *Метод наименьших квадратов*. Минимизируется сумма квадратов отклонений выборочных данных от определяемой оценки.

Пример 1.

$$U=\sum_{i=1}^{n}\left(X_i-\theta\right)^2;\ \left.\frac{dU}{d\theta}\right|_{\theta=\hat{\theta}_n}=-2\sum_{i=1}^{n}\left(X_i-\hat{\theta}_n\right)=0\therefore\ \hat{\theta}_n=\frac{1}{n}\sum_{i=1}^{n}X_i=\hat{\bar{X}}_n=0$$

(оценка для генеральной средней). Искомая оценка есть ***средняя выборочная***.

(3) *Метод максимального правдоподобия*. Пусть $p(x,\theta)$ плотность распределения генеральной совокупности. Тогда для возвратной случайной выборки, рассматриваемой по схеме независимых испытаний, имеет место

$$p\left(x_{(1)},\ \ldots,\ x_{(n)},\ \theta\right)=p\left(x_{(1)},\ \theta\right)\cdot p\left(x_{(2)},\ \theta\right)\cdot\ldots\cdot p\left(x_{(n)},\ \theta\right),$$

где $x_{(1)},\ \ldots,\ x_{(n)}$ — есть переменные, а θ — заданный постоянный параметр.

Определение 1. Функция вида

$$p\left(x_{(1)},\ \ldots,\ x_{(n)},\theta\right)=p\left(x_{(1)},\theta\right)\cdot p\left(x_{(2)},\theta\right)\cdot\ldots\cdot p\left(x_{(n)},\theta\right)=$$
$$=p_{x_1}(\theta)\cdot p_{x_2}(\theta)\cdot\ldots\cdot p_{x_n}(\theta),$$

где $x_1,\geq,x_n$ фиксированы, а θ — переменная, называется ***функцией правдоподобия выборки***.

Пусть

$$L\left(x_1,\ \ldots,\ x_n,\theta\right)\equiv\ln p\left(x_{(1)},\ \ldots,\ x_{(n)},\theta\right)=\sum_{i=1}^{n}\ln p\left(x_{(i)},\theta\right)=\sum_{i=1}^{n}\ln p_{x_i}(\theta).$$

Тогда $\hat{\theta}$ ищется из условия: $\left.\dfrac{dL(\ )}{d\theta}\right|_{\theta=\hat{\theta}}=0=\sum\limits_{i=1}^{n}\dfrac{1}{p_{x_i}(\theta)}\cdot\dfrac{dp_{x_i}(\theta)}{d\theta}.$

Замечание 1. В случае оценки нескольких параметров $\theta_1,\ \ldots,\ \theta_n$ очевидно необходимым условием максимума L является $\dfrac{\partial L}{\partial\theta_i}=0;\ i=1,\ \ldots,\ n.$

Пример 2. Генеральная совокупность с нормальным распределением

$$p_N(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\bar{x})^2}{2\sigma^2}},\ \text{где}\ \theta_1=\bar{X},\ \theta_2=\sigma^2.$$

Функция правдоподобия для возвратной выборки:

$$p_N(x_1,\ \ldots,\ x_n,\ \theta_1,\ \theta_2)=\frac{1}{(2\pi\theta_2)^{n/2}}\cdot e^{-\frac{\sum(x_i-\theta_1)^2}{2\theta_2}}.$$

$$L(\ )=-\frac{n}{2}(\ln 2\pi+\ln\theta_2)-\frac{\sum(x_i-\theta_1)^2}{2\theta_2}.$$

а) $\left.\dfrac{\partial L}{\partial\theta_1}\right|_{\theta_1=\hat{\theta}_1}=\dfrac{1}{2\theta_2}\sum 2\left(X_i-\hat{\theta}_1\right)=0\Rightarrow\hat{\theta}_1=\dfrac{1}{n}\sum X_i\ \therefore\ \hat{\theta}_1=\hat{\bar{\bar{X}}}$

выборочная средняя.

б) $\left.\dfrac{\partial L}{\partial\theta_2}\right|_{\theta_2=\hat{\theta}_2,\ \theta_1=\hat{\theta}_1}=-\dfrac{n}{2\hat{\theta}_2}+\dfrac{1}{2\hat{\theta}_2^2}\sum\left(X_i-\hat{\theta}_1\right)^2=0\Rightarrow\hat{\theta}_2=$

$=\dfrac{1}{n}\sum\left(X_i-\hat{\theta}_1\right)^2=\dfrac{1}{n}\sum\left(X_i-\hat{\bar{\bar{X}}}\right)^2=s_n^2$

выборочная дисперсия.

14.2. Оценка доли признака

При оценке доли признака A имеем генеральное распределение с долей p. Возвратная выборка объема n снова рассматривается как n повторных независимых испытаний, и вероятность получить выборку с относительной частотой признака $\hat{p}=\dfrac{m}{n}$ можно определить по формуле:

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

Пусть p — оцениваемый параметр, тогда функция правдоподобия:

$$p\left(\frac{m}{n},\ \theta\right)=C_n^m\theta^m\left(1-\theta\right)^{n-m} \text{ и } L\left(\frac{m}{n},\theta\right)=\ln C_n^m+m\ln\theta+\left(n-m\right)\ln\left(1-\theta\right)$$

$$\left.\frac{\partial L}{\partial\theta}\right|_{\theta=\hat{\theta}}=\frac{m}{\hat{\theta}}-\frac{n-m}{1-\hat{\theta}}=0\Rightarrow\hat{\theta}=\frac{m}{n} \text{ и } \hat{p}=\frac{m}{n}.$$

Эта точечная оценка проверяется на соответствующие требования:

1) Состоятельность оценки $\hat{p}=\dfrac{m}{n}$ следует из закона больших чисел (ЗБЧ) в форме Бернулли.
2) Несмещенность оценки: $E\hat{p}=E\left(\dfrac{m}{n}\right)=p.$
3) Достаточность оценки следует из условной вероятности, определяемой как отношение вероятности одной данной выборки $p^m\cdot q^{n-m}$ к вероятности любой выборки $C_n^m p^m\cdot q^{n-m}$ и равной $\dfrac{1}{C_n^m}$ (не зависит от p).
4) Можно доказать, что $\hat{p}$ есть эффективная оценка.

   При нахождении *предельной ошибки выборки* используется интервальная оценка

$$P\left(\hat{p}-\varepsilon_1<p<\hat{p}+\varepsilon_2\right)=1-\alpha;\ \hat{p}-\varepsilon_1\equiv p_1 \text{ и } \hat{p}+\varepsilon_2\equiv p_2.$$

При выборке объема $n<20$ используется метод расчета по биномиальному закону, существуют таблицы для определения p_1 и p_2 по данным α, n, $\hat{p}$.

При $np\geq10$, т. е. при $n\geq20$ используется асимптотическая формула Лапласа, а при малых p асимптотическая формула Пуассона:

$$P\left(\left|\hat{p}-p\right|<\varepsilon_\alpha\right)=2\Phi\left(z_\alpha\right)=1-\alpha \text{ при } \varepsilon_\alpha=z_\alpha\sigma\left(\hat{p}\right);$$

$$\sigma\left(\hat{p}\right)=\sqrt{\frac{pq}{n}}\approx\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}.$$

14.3. Точечные оценки для средней и дисперсии генеральной совокупности

Пусть $\bar{X}$ и σ^2 — среднее и дисперсия генеральной совокупности. Выборка рассматривается как система n случайных величин, одинаково распределенных и совпадающих с генеральным:

$$EX_i = \bar{X} \text{ и } DX_i = \sigma^2.$$

Возвратная выборка объема n, где X_i — независимые случайные величины.

(1) Для оценки генеральной *средней* $\bar{X}$ выберем статистику $\theta = \hat{\bar{X}}_n$. Пусть статистика $\theta = \hat{\bar{X}}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$ — выборочная средняя. Тогда для нее:

$$E\hat{\bar{X}}_n = \frac{1}{n}E\left(\sum X_i\right) = \frac{1}{n}\sum EX_i = \frac{n\bar{X}}{n} = \bar{X},$$

$$D\hat{\bar{X}}_n = \frac{1}{n^2}D\left(\sum X_i\right) = \frac{1}{n^2}\sum DX_i = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

В силу ЗБЧ (теорема Чебышева) $\hat{\bar{X}}_n \xrightarrow{\text{Вер.}} \bar{X}$ оценка *состоятельна*, а *несмещенность* следует из предыдущего равенства. Для нормального распределения генеральной совокупности эта оценка является *эффективной*, что проверяется через неравенство Рао-Крамера. Таким образом $\boxed{\bar{X} \approx \hat{\bar{X}}_n}$.

(2) Для оценки генеральной *дисперсии* выберем статистику:

$\theta = s_n^2 \equiv \frac{1}{n}\sum\left(X_i - \hat{\bar{X}}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \hat{\bar{X}}_n^2$ — выборочная дисперсия.

С учетом $\hat{\bar{X}}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ и $\hat{\bar{X}}_n^2 = \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j\right)$

имеем $s_n^2 = \frac{n-1}{n^2}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\sum_{i \neq j} X_i X_j$.

С учетом свойства дисперсии $DX=D(X+C)$ из последнего равенства справа из всех X_i и X_j можно вычесть $\bar{X}$. И тогда, проверяя статистику на смещенность, получим

$$Es_n^2 = \frac{n-1}{n^2}\sum_{i=1}^{n} E\left(X_i - \bar{X}\right)^2 - \frac{1}{n^2}\sum_{\substack{i,j\\ i\neq j}} E\left(X_i - \bar{X}\right)\left(X_j - \bar{X}\right) =$$

$$= \frac{n-1}{n^2}\cdot n\sigma^2 = \frac{n-1}{n}\sigma^2.$$

Статистика оказалась смещенной (смещение $a = -\frac{1}{n}\sigma^2$). Обычно вводится «скорректированная» выборочная дисперсия:

$$\boxed{s^2 \equiv \frac{n}{n-1}s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \hat{\hat{X}}_n\right)^2}.$$

Эта оценка не смещена, и в силу ЗБЧ состоятельна. Таким образом, $\boxed{\sigma^2 \approx s^2}$.

Замечание 2. При больших n: $\frac{n}{n-1} \approx 1 \Rightarrow s^2 \approx s_n^2$.

Безвозвратная выборка. Точечная оценка среднего не изменится, а для дисперсии заменится на $\boxed{s'^2 = \frac{N-1}{N}s^2}$, где N — объем генеральной совокупности.

$$\boxed{D'\left(\hat{\hat{X}}_n\right) = \frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}} \text{ и } \boxed{D\left(s^2\right) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right)}.$$

15. Интервальные оценки параметров

15.1. Оценки средней и дисперсии нормально распределенной генеральной совокупности

А. *Оценка средней при известной генеральной дисперсии* σ^2.

Генеральная совокупность: $p(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{X})^2}{2\sigma^2}}$.

1) Оценка $\hat{\hat{X}}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ есть случайная величина

$$\hat{\hat{X}}_n \in N\left(\bar{X}, \frac{\sigma^2}{n}\right).$$

2) Тогда $\delta = \hat{\hat{X}}_n - \bar{X}$ тоже случайная величина, а потому

3) $P\left(\left|\hat{\hat{X}}_n - \bar{X}\right| < \varepsilon\right) = P\left(\hat{\hat{X}}_n - \varepsilon < \bar{X} < \hat{\hat{X}}_n + \varepsilon\right) = 2\Phi(z)$,

где $z = \dfrac{\varepsilon}{\sigma\left(\hat{\hat{X}}_n\right)} = \dfrac{\varepsilon \cdot \sqrt{n}}{\sigma}$.

Случаи:

I. Пусть задана доверительная вероятность P. Найти ε и доверительный интервал.

Решение:

$$P = 1 - \alpha = 2\Phi(z) \Rightarrow z_\alpha = \Phi^{-1}\left(\frac{1-\alpha}{2}\right)$$

(по таблице функций Лапласа); $\varepsilon_\alpha = z_\alpha \dfrac{\sigma}{\sqrt{n}}$,

и тогда доверительный интервал есть $\left(\hat{\hat{X}}_n - \varepsilon_\alpha, \hat{\hat{X}}_n + \varepsilon_\alpha\right)$.

II. Задан объем выборки n и предельная ошибка ε. Найти доверительную вероятность P.

Решение: $z = \dfrac{\varepsilon\sqrt{n}}{\sigma} \Rightarrow \Phi(z) \Rightarrow P = 2\Phi(z)$.

III. Заданы P и ε, найти объем выборки n.

Решение: $P = 2\Phi(z) \Rightarrow z = \Phi^{-1}\left(\dfrac{P}{2}\right) \Rightarrow n = \dfrac{z^2\sigma^2}{\varepsilon^2}$.

B. *Оценка средней при неизвестной генеральной дисперсии* σ^2.

Далее потребуются два производных от нормального распределения случайной величины:

(1) — *Распределение Пирсона* (χ^2-распределение) случайной величины $\chi^2 = \sum_{i=1}^{n} \dfrac{\left(X_i - \bar{X}\right)^2}{\sigma^2}$, где X_i — одинаково нормально распределенные случайные величины $X_i \in N\left(\bar{X}, \sigma^2\right)$. Плотность распределения случайной величины χ^2 определяется как плотность функции случайной величины и имеет вид

$$p_{\chi^2}(x,n) = \begin{cases} K e^{-\frac{x}{2}} \cdot x^{\frac{n}{2}-1}, \ x > 0 \\ 0, \ x \le 0 \end{cases}, \text{ где } K = 1 : \int_0^{\infty} e^{-\frac{y}{2}} \cdot y^{\frac{n}{2}-1} dy.$$

Если X_i удовлетворяют ℓ линейным зависимостям, то

$p_{\chi^2}(x,\nu) = K e^{-\frac{x}{2}} \cdot x^{\frac{\nu}{2}-1}$, где $\nu = n - l$ — число степеней свободы случайной величины X_i.

$E\chi^2 = \nu$; $D\chi^2 = 2\nu$. При $\nu \ge 30$ χ^2 стремится к нормальному.

(2) — *Распределение Стьюдента* (t-распределение). Пусть $Z \in N(0, 1)$, U — случайная величина из χ^2-распределения с ν степенями свободы.

Величина $T = \dfrac{Z\sqrt{\nu}}{U}$ подчиняется t-распределению с плотностью

$$p_s(t,\nu) = B\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \text{ где } B = 1 : \int_{-\infty}^{\infty} \left(1 + \frac{q^2}{\nu}\right)^{-\frac{\nu+1}{2}} dq.$$

$ET = 0$; $D(T) = \frac{\nu}{\nu - 2}$. При $\nu \geq 20$ t-распределение стремится к нормальному.

Тогда:

$$P(T < t) = S(t,\nu) = B\int\limits_{-\infty}^{t}\left(1 + \frac{r^2}{\nu}\right)^{-\frac{\nu+1}{2}}dr \approx 0{,}5 + \Phi(t).$$

Введем статистику $T \equiv \frac{\hat{\bar{X}}_n - \bar{X}}{s}\sqrt{n}$. Это есть статистика Стьюдента со степенью свободы $\nu = n - 1$.

Используя функцию распределения $S(t,\,n) = P(T < t)$ аналогично формуле Лапласа $P\left\{\frac{\left|\hat{\bar{X}}_n - \bar{X}\right|}{s}\sqrt{n} < t\right\} = 2S(t,\nu) - 1 = 1 - \alpha$ или

$$P\left\{\hat{\bar{X}}_n - \varepsilon < \bar{X} < \hat{\bar{X}}_n + \varepsilon\right\} = 2S(t,\nu) - 1, \text{ где } \varepsilon = \frac{ts}{\sqrt{n}}.$$

Тогда $S(t,\nu)^{-1} = 1 - \frac{\alpha}{2} \Rightarrow t_\alpha = S^{-1}\left(1 - \frac{\alpha}{2}, \nu\right)$. Далее аналогично п. А, случаи I–III.

В случае III используется схема последовательных приближений. Вначале находят n_1 при $\sigma^2 = s^2$. По n_1 и $\nu_1 = n_1 - 1$, заданному $P = 1 - \alpha$, определяют t_1 и вычисляют $n_2 = \frac{t_1^2 s^2}{\varepsilon^2}$ и т. д.

C. *Оценки дисперсии*. Интервальная оценка дисперсии.

Введем статистику $\chi^2 = \frac{s^2}{\sigma^2}(n-1)$. Она подчиняется χ^2-распределению с $\nu = n - 1$ степенями свободы.

Тогда $P\left(u_1 < \chi^2 < u_2\right) = F_{\chi^2}\left(u_2, \nu\right) - F_{\chi^2}\left(u_1, \nu\right)$. После преобразования интервала слева имеем

$$P\left\{\frac{s^2(n-1)}{u_2} < \sigma^2 < \frac{s^2(n-1)}{u_1}\right\} = F_{\chi^2}\left(u_2, \nu\right) - F_{\chi^2}\left(u_1, \nu\right).$$

Если допустить:

$$\left.\begin{aligned} P\left(\chi^2 < u_1\right) &= F_{\chi^2}\left(u_1, \nu\right) = \frac{\alpha}{2} \\ P\left(\chi^2 < u_2\right) &= F_{\chi^2}\left(u_2, \nu\right) = 1 - \frac{\alpha}{2} \end{aligned}\right\}, \text{ то } \overbrace{\left(\frac{s^2(n-1)}{u_2}; \frac{s^2(n-1)}{u_1}\right)}^{\textit{доверительный интервал}}.$$

15.2. Приближенный метод интервальной оценки генеральной средней

А. *Оценка средних для больших выборок*. При $n \geq 20$ по теореме Ляпунова практически среднеарифметическая распределена нормально, т. е. $\hat{\bar{\bar{X}}}_n \in N\left(\bar{X}, \frac{\sigma}{\sqrt{n}}\right)$. Но σ неизвестна. Если заменить ее на оценку s или s_n, то для стандартного уклонения оценки среднего имеем $\sigma\left(\hat{\bar{\bar{X}}}_n\right) = \frac{s_n}{\sqrt{n}}$ или $\sigma\left(\hat{\bar{\bar{X}}}_n\right) = \frac{s}{\sqrt{n}}$. По аналогии с п. 14.1.А предельная ошибка выборки $\varepsilon_\alpha = z_\alpha \sigma\left(\hat{\bar{\bar{X}}}\right)$, а доверительный интервал для средней $\bar{X}$ есть $\left(\hat{\bar{\bar{X}}}_n - \varepsilon_\alpha < \bar{X} < \hat{\bar{\bar{X}}}_n + \varepsilon_\alpha\right)$.

Далее схема прежняя (п. 14.1.А), как при известной дисперсии генеральной совокупности.

В. *Оценка средней по малой выборке* ($n < 10 \div 20$). При $n < 10 \div 20$ центральная предельная теорема неприменима и замена неизвестной σ^2 на s^2 неправомочна. Для малой выборки используется лишь выборочная средняя $\hat{\bar{\bar{X}}}_n$ как точечная оценка с проверкой на *условную эффективность*.

Пусть $\theta = \sum_{i=1}^{n} a_i X_i$, где $\sum_{i=1}^{n} a_i = 1$ и $a_i = const$, θ — *линейная оценка*:

$E\theta = \sum a_i E\text{X}_i = \bar{X} \cdot \sum a_i = \bar{X}$, т. е. θ — несмещенная.

$D\theta = \sum {a_i}^2 D\text{X}_i = \sigma^2 \sum_{i=1}^{n} {a_i}^2$. Минимум $D\theta$ достигается при

$a_1 = a_2 = \ldots = a_n = \frac{1}{n}$. Таким образом, $\widehat{\widehat{X}}_n$ обладает минимальной дисперсией из линейных оценок θ, т. е. она эффективна.

15.3. Статистические оценки при многоступенчатом отборе

Пусть генеральная совокупность разбита на ℓ непересекающихся групп (серий, гнезд) численностью $N_1, \ldots, N_i, \ldots, N_\ell$ элементов; $\bar{X}_i$ — групповая средняя; D_i — групповая дисперсия. Известно, что для всей совокупности:

$$\begin{cases} \bar{X} = \frac{1}{N} \sum_{i=1}^{\ell} \bar{X}_i N_i;\ DX = \bar{D}_{\text{а. гр.}} + D_{\text{м. гр.}},\ \text{где} \\ \bar{D}_{\text{а. гр.}} = \frac{1}{N} \sum_{i=1}^{\ell} DX_i \cdot N_i \ \text{— средняя групповых дисперсий} \\ D_{\text{м. гр.}} = \frac{1}{N} \sum_{i=1}^{\ell} \left(\bar{X}_i - \bar{X}\right)^2 N_i \ \text{— дисперсия групповых средних} \end{cases}$$

А. *Типическая (типологическая, стратифицированная) выборка*. (В группах более однородные элементы.) Выборка состоит из: n_1 — элементов из 1-й группы; n_2 — элементов из 2-й группы; $\ldots n_\ell$ — элементов ℓ-й группы. Тогда $\widehat{\widehat{X}}_i$ — выборочная групповая средняя, а $\widehat{D}_i$ — выборочная групповая дисперсия.

Пусть для всей выборки находится $\widehat{\widehat{X}}$ — выборочная средняя.

Оценка: $\boxed{\widehat{\widehat{X}} = \frac{1}{n} \sum \widehat{\widehat{X}}_i n_i}$. При $\frac{n_i}{n} = \frac{N_i}{N}$ оценка не смещена $E\widehat{\widehat{X}} = \bar{X}$.

Ее дисперсия $D\left(\hat{\bar{X}}\right)_{\text{тип}} = \sum_i \frac{n_i^2}{n^2} D\left(\hat{\bar{X}}_i\right)$, но $D\hat{\bar{X}}_i = \frac{\sigma_i^2}{n_i} = \frac{D_i}{n_i}$.

Тогда $D\left(\hat{\bar{X}}\right)_{\text{тип}} = \frac{1}{n} \sum D_i \hat{\bar{X}}_i \frac{N_i}{N} = \frac{\bar{D}_{\text{а. гр.}}}{n}$.

Замечание. При *обычной случайной* выборке:

$$D\left(\hat{\bar{X}}\right) = \frac{D}{n} = \frac{D_{\text{м. гр.}}}{n} + \frac{\bar{D}_{\text{а. гр.}}}{n} = D\left(\hat{\bar{X}}\right)_{\text{тип}} + \frac{D_{\text{м. гр.}}}{n}.$$

Интервальная оценка находится с учетом дисперсии, полученной выборочной средней.

$$P\left(\left|\hat{\bar{X}} - \bar{X}\right| < \varepsilon\right) = 2\Phi(z), \text{ где } z = \frac{\varepsilon}{\sigma\left(\hat{\bar{X}}\right)} = \frac{\varepsilon\sqrt{n}}{\sigma_{\text{а. гр.}}},$$

но $\sigma^2_{a.\text{ гр.}} = \bar{D}_{a.\text{ гр}} = \frac{1}{N}\sum_i D_i N_i = \frac{1}{n}\sum_i D_i n_i \approx \frac{1}{n}\sum_i s_i^2 n_i$ и т. д.

В. *Серийная выборка* (генеральная совокупность разбита на ℓ равновеликих серий). Отобрано k групп. Выборочная совокупность состоит из $n = k\frac{N}{\ell}$. Случайный отбор производится группами. $\hat{\bar{X}} = \frac{1}{k}\sum \bar{X}'_j$. Оценка несмещенная, но с большей ошибкой, чем простая случайная выборка, и тем более типическая.

16. Введение в теорию проверки статистических гипотез

16.1. Общая постановка задачи

Определение 1. Любое проверяемое по выборочным данным предположение о виде или значениях параметров распределения генеральной совокупности называют ***статистической гипотезой***.

Суть проверки статистической гипотезы заключается в том, чтобы установить, согласуются ли данные наблюдения с выдвинутой гипотезой.

Примеры:

1. $H_o: X \in N(a, \sigma)$, где $a = a_o$, σ — известна.
2. $H_o: X \in N(a, \sigma)$, где a — неизвестна, $\sigma = \sigma_o$.

Определение 2. Подлежащая проверке гипотеза называется ***основной*** или ***нулевой*** (H_0), а логически отрицающая ее гипотеза называется ***конкурирующей***, или альтернативной (H_a) или (H_1).

Определение 3. Гипотеза называется ***параметрической***, если в ней содержится утверждение о параметрах или числовых характеристиках генерального распределения, и ***непараметрической***, если она содержит утверждение обо всем распределении.

Определение 4. Параметрическая гипотеза называется ***простой***, если в ней речь идет ровно об одном значении одномерного или многомерного параметра (например, H_0: $x = a$), в противном случае ***сложной***.

При проверке гипотезы выборочные данные могут согласовываться с основной гипотезой или ей противоречить, поэтому всегда есть риск принять ложное решение. Из четырех возможных решений два могут оказаться ложными:

1. Гипотеза H_0 верна, и она принимается — верное решение.
2. Гипотеза H_0 не верна, и она отвергается — верное решение.
3. Гипотеза H_0 верна, но она отвергается — ошибка 1 рода.
4. Гипотеза H_0 не верна, но она принимается — ошибка 2 рода.

16.2. Критерий проверки. Критическая область

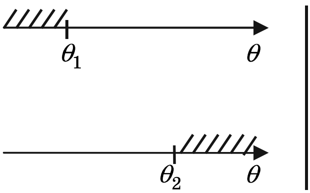
Определение 5. ***Статистическим критерием*** называют однозначно определенное правило, руководствуясь которым проверяемую гипотезу отклоняют или не отклоняют. Основой критерия является специально составленная выборочная характеристика (статистика) $\theta = f(x_1, \geq, x_n)$, точное или приближенное распределение которой известно.

Определение 6. Пусть область значений θ как функции состоит из двух подмножеств, в одном из которых гипотеза H_0 отклоняется, а в другом не отклоняется. Тогда первое называется ***критической областью*** w_0, второе — ***областью принятия гипотезы*** (допустимой областью), а их граница — ***критической точкой (критическими точками)***.

Замечание 1. Случайная величина θ может быть одномерной и многомерной.

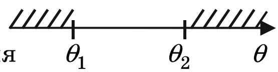
Примеры критических точек и критических областей:

1.



— односторонняя критическая область

2.



двусторонняя критическая область

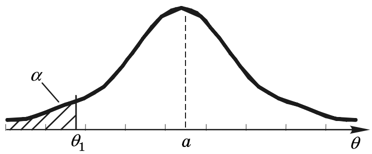
Определение 7. ***Уровнем значимости*** α критерия называют вероятность совершить ошибку 1-го рода, т. е. $P\{\theta \in w_0\} = \alpha$.

Замечание 2. Обычно принимают $\alpha = 0{,}05$ или $0{,}01$.

Замечание 3. С уменьшением α увеличивается вероятность ошибки 2-го рода (β), а при $\alpha = 0$ гипотеза всегда принимается. Следовательно, α и β являются конкурирующими.

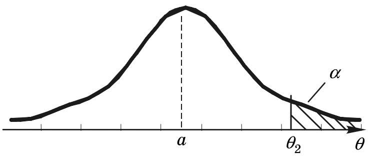
Примеры:

1)

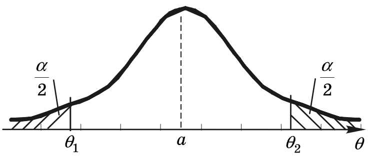


$$P\{\theta < \theta_1\} = \alpha$$
$$H_0 : \bar{x} = a$$
$$H_1 : \bar{x} < a$$

2)



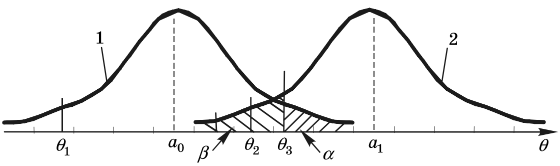
$$P\{\theta > \theta_2\} = \alpha$$
$$H_0 : \bar{x} = a$$
$$H_1 : \bar{x} > a$$

3) 

$$P\begin{Bmatrix}\theta \ge \theta_2 \\ \theta < \theta_1\end{Bmatrix} = \alpha$$

$H_0 : \bar{x} = a$

$H_1 : \bar{x} \neq a$

4) 

$a) \; H_0 : \bar{x} = a_0$

$\quad H_1 : \bar{x} \neq a_0$

$b) \; H_0 : \bar{x} = a_0$

$\quad H_1 : \bar{x} = a_1$

$$\beta = P\left\{\bar{x} \le \theta_2\right\}$$

При $a_0 \to a_1$ уменьшение α ведет к росту β (при близких a_0 и a_1 рост очень быстрый). График зависимости $\beta(\alpha)$ называют ***кривыми эффективности критерия***.

Определение 8. Вероятность принять конкурирующую гипотезу, когда она верна, называется ***мощностью критерия***, которая равна $(1-\beta)$.

Определение 9. Критерий называется ***наиболее мощным***, если для всех критериев с уровнем α он обладает наибольшей мощностью: $\max\limits_{\theta} P\left\{\theta \in \omega_0 \middle| H_1\right\}$

$$P\left\{\theta \in \omega_0 \middle| H_0\right\} = \alpha \text{ и } P\left\{\theta \in \omega_0 \middle| H_1\right\} = 1 - \beta = \max.$$

16.3. Общая схема проверки гипотез

1. Формулируются проверяемая H_0 и альтернативная H_1 гипотезы.
2. Выбирается статистика θ (определяется по выборочным данным). Можно предлагать различные статистики, удовлетворяющие требованиям:

- должен быть известен закон распределения θ в зависимости от гипотезы;
- статистика должна быть **несмещенной, состоятельной** и **эффективной.**

3. Формулируется правило проверки, определяется объем выборки по заданным α, β (или из условия минимизации β по α и n).
4. Выбирается одно- или двухсторонняя проверка в зависимости от гипотезы и ее альтернативы.
5. Определяются критические точки и области по величине α.
6. Производится выборка намеченного объема n, по которой вычисляется выборочное значение $\hat{\theta}$, принимается решение о ее принадлежности критической области.

Замечание 4. В критериях наиболее часто используются χ^2-статистика Пирсона, t-статистика Стьюдента, F-статистика Фишера, z-статистика (стандартная нормальная).

17. Проверка параметрических гипотез

17.1. Проверка гипотез относительно доли признака

1. Сравнение доли признака с нормативом.
   Пусть p — доля некоторого признака в генеральной совокупности, a — нормативная величина.

   а) H_0: $p = \text{a}$; H_1: $p \neq \text{a}$. За статистику принимают $\theta \equiv \dfrac{m}{n}$ — частость признака в выборке. Очевидно, θ распределен по биномиальному закону, а при $n \to \infty$ по асимптотически нормальному. Пусть α — уровень значимости.

   Тогда $P\left\{\left|\dfrac{m}{n} - a\right| \leq \overbrace{z_{\alpha/2}\sigma\left(\dfrac{m}{n}\right)}^{\varepsilon}\right\} = 2\Phi(z) = 1 - \alpha.$

   Откуда по таблицам Лапласа находим $z_{\alpha/2}$.

В силу биномиального закона принимаем

$$\sigma\left(\frac{m}{n}\right)=\sqrt{\frac{a(1-a)}{n}}\ \therefore.$$

Доверительные границы для $\frac{m}{n}$ и критической точки следуют из неравенства:

$$\theta_1=a-z_{\alpha/2}\sqrt{\frac{a(1-a)}{n}} \text{ и } \theta_2=a+z_{\alpha/2}\sqrt{\frac{a(1-a)}{n}}.$$

Если выборочная частость $\hat{\theta}\in(\theta_1,\theta_2)$, то H_0 принимается (не отвергается).

б) H_0: $p=a$; H_1: $p>a \therefore P\left\{\frac{m}{n}>\theta_2\right\}=0{,}5-\Phi(z)=\alpha$, по таблице находится z_α, а затем $\theta_2=\mathrm{a}+z_a\sqrt{\frac{\mathrm{a}(1-\mathrm{a})}{n}}$; H_0 отклоняется, если выборочная частость $\hat{\theta}=\frac{m}{n}>\theta_2$.

2. Сравнение доли признака в двух совокупностях.

$\frac{m_1}{n_1}$, $\frac{m_2}{n_2}$ — частости одного и того же признака в двух совокупностях из n_1 и n_2 единиц.

Гипотезы H_0:{обе совокупности} из одной генеральной совокупности с долей p. H_1: совокупности не из одной генеральной совокупности.

а) Большие выборки. n_1, $n_2>30 \therefore$ Распределение частостей близко к нормальному. Они не смещены $E\frac{m_1}{n_1}=E\frac{m_2}{n_2}=p$

и имеют дисперсии: $\sigma^2\left(\frac{m_1}{n_1}\right)=\frac{p(1-p)}{n_1}$ и $\sigma^2\left(\frac{m_2}{n_2}\right)=\frac{p(1-p)}{n_2}$.

Принимаем статистику: $\theta \equiv \frac{m_1}{n_1} - \frac{m_2}{n_2} \therefore \theta \in N\left(0,\ \sigma^2(\theta)\right)$, где

$$\sigma^2(\theta) = \sigma^2\left(\frac{m_1}{n_1}\right) + \sigma^2\left(\frac{m_2}{n_2}\right) = p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

Пусть $P\{|\theta| < z_{\alpha/2}\sigma(\theta)\} = 2\Phi(z) = 1-\alpha$.

Откуда $\theta_{1,2} = \mp z_{\alpha/2}\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, где оценка p произведена по двум выборкам $p = \frac{m_1 + m_2}{n_1 + n_2}$. Тогда имеем $\hat{\theta} = \frac{m_1}{n_1} - \frac{m_2}{n_2}$ и правило проверки: $\hat{\theta} \in (\theta_1, \theta_2) \Rightarrow H_0$ не отвергается.

б) Малые выборки. (Используется χ^2-статистика Пирсона).

17.2. Проверка гипотез относительно средней

1. Сравнение средней с нормативом.

$H_0 : \bar{X} = a$ (т. е. генеральная средняя равна заданному числу а).
Два случая:

а) Большая выборка, и генеральная дисперсия σ^2 известна. Тогда в силу центральной предельной теоремы выборочная средняя $\hat{\bar{\bar{X}}}_n$ распределена по нормальному закону.

Здесь критерий проверки $z = \frac{\hat{\bar{\bar{X}}}_n - a}{\sigma\left(\hat{\bar{\bar{X}}}_n\right)}$ (нормированное отклонение). Тогда $z \in N(0,1)$. Задавшись уровнем значимости α, найдем для $H_1 : \hat{\bar{\bar{X}}} \neq$ а из $P\left\{|z| < z_{\alpha/2}\right\} = 2\Phi\left(z_{\alpha/2}\right) = 1-\alpha$ значение

$z_{\alpha/2} = \Phi^{-1}\left(\frac{1-\alpha}{2}\right)$. Тогда критические точки $z_1 = -z_{\alpha/2}$; $z_2 = +z_{\alpha/2}$.

Следовательно, если значение статистики $\hat{z} \in \left(-z_{\alpha/2}, z_{\alpha/2}\right)$, то H_0 не отклоняется.

Для альтернативных гипотез $H_1: \bar{X} \gtrless a$ производится односторонняя проверка при помощи z_α, определяемого из уравнения: $\alpha = P\{/z/ < z_\alpha\} = 0{,}5 + \Phi(z_\alpha)$.

Для альтернативы $H_1: \bar{X} = a_1$ можно определить и мощность критерия $1-\beta$. Пусть $a_1 > a$, тогда β определяется из уравнения $\beta = P\{|z/ < z_\alpha\}$. Перейдя от z к нормированному отклонению от нового центра a_1, имеем:

$$z' = \frac{\hat{\bar{X}}_n - a_1}{\sigma\left(\hat{\bar{X}}_n\right)} = \frac{\hat{\bar{X}}_n - a}{\sigma\left(\hat{\bar{X}}_n\right)} + \frac{a - a_1}{\sigma\left(\hat{\bar{X}}_n\right)} = z + \lambda,\ \lambda \equiv \frac{a - a_1}{\sigma\left(\hat{\bar{X}}_n\right)}.$$

Тогда:

$$\beta = P\left\{|z'| < z_{\alpha/2}\right\} = P\left\{\lambda - z_{\alpha/2} < z < \lambda + z_{\alpha/2}\right\} = \Phi\left(\lambda + z_{\alpha/2}\right) - \Phi\left(\lambda - z_{\alpha/2}\right).$$

б) Малая выборка, и $\sigma\left(\hat{\bar{X}}_n\right)$ неизвестна. При неизвестной σ для $n > 30$ ее можно заменить на выборочную несмещенную $s = \sqrt{\frac{1}{n-1}\sum\left(X_i - \hat{\bar{X}}_n\right)^2}$. Если $n < 30$, то удобнее использовать t-статистику Стьюдента с $\nu = n - 1$ степенями свободы: $t = \frac{\hat{\bar{X}}_n - a}{s}\sqrt{n}$, а далее аналогично схеме а), заменяя в выражении для вероятности функцию Лапласа функцией Стьюдента (п. 14.1).

2. Сравнение средних двух совокупностей.

$$\left.\frac{\text{Первая совокупность } \bar{X},\ \sigma_x}{\text{Вторая совокупность } \bar{Y},\ \sigma_y}\right\}\ H_0:\ \bar{X} = \bar{Y}\ \left\{\frac{\text{Выборки } n_1,\ \hat{\bar{X}}_{n_1},\ s_x}{\text{Выборки } n_2,\ \hat{\bar{Y}}_{n_2},\ s_y}\right.$$

Используется статистика $\boxed{\theta = \dfrac{\hat{\hat{X}}_{n_1} - \hat{\hat{Y}}_{n_2}}{\sigma\left(\hat{\hat{X}}_{n_1} - \hat{\hat{Y}}_{n_2}\right)}}$. (В дальнейшем индексы при выборочных средних опускаем.) В силу несмещенности $\hat{\hat{X}}$ и $\hat{\hat{Y}} \to E$ имеем $E\theta = 0$.

В силу независимости выборок n_1 и n_2 имеем

$$\sigma^2\left(\hat{\hat{X}} - \hat{\hat{Y}}\right) = \sigma^2\left(\hat{\hat{X}}\right) + \sigma^2\left(\hat{\hat{Y}}\right) = \frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}.$$

Предположение: $\sigma_x^2 = \sigma_y^2 = \sigma^2$.

Тогда $\sigma^2\left(\hat{\hat{X}} - \hat{\hat{Y}}\right) = \sigma^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)$, а статистика $\theta = \dfrac{\hat{\hat{X}} - \hat{\hat{Y}}}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$.

При $n > 20 \div 30$ $\hat{\hat{X}}$, $\hat{\hat{Y}}$ и θ распределены приблизительно нормально.

Полагая $\sigma^2 \approx s^2 = \dfrac{\sum\left(X_i - \hat{\hat{X}}\right)^2 + \sum\left(Y_i - \hat{\hat{Y}}\right)^2}{n_1 + n_2 - 2} = \dfrac{n_1 s_x^2 + n_2 s_y^2}{n_1 + n_2 - 2}$,

будем иметь нормально распределенную статистику

$$\boxed{\hat{\theta} = \frac{\hat{\hat{X}} - \hat{\hat{Y}}}{\sqrt{n_1 s_x^2 + n_2 s_y^2}} \cdot \frac{\sqrt{n_1 + n_2 - 2}}{\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}}.$$

Далее проверка ведется обычным образом с применением функций Лапласа или Стьюдента.

17.3. Сравнение дисперсий двух нормальных совокупностей

Первая совокупность нормальная, σ_1^2

Вторая совокупность – нормальная, σ_2^2

$H_0: \sigma_1^2 = \sigma_2^2$

Выборочная дисперсия s_1^2 при объеме n_1

Выборочная дисперсия s_2^2 при объеме n_2

Используется статистика $F = \dfrac{s_1^2}{s_2^2}$, где $s_1^2 > s_2^2$ (Фишер—Снедекор).

Распределение ФишераСнедекора. Пусть U и V — независимые случайные величины, имеющие χ^2-распределение со степенями свободы ν_1 и n_2, тогда $F = \dfrac{U}{\nu_1} : \dfrac{V}{\nu_2}$ зависит от двух параметров ν_1 и ν_2 с плотностью $\boxed{p_F(x, \nu_1, \nu_2) = \begin{cases} C \cdot x^{\frac{\nu_1}{2} - 1} \left(1 + \dfrac{\nu_1}{\nu_2} x\right)^{-\frac{\nu_1 + \nu_2}{2}} \\ 0, \quad x < 0 \end{cases}}$. При ν_1, $\nu_2 >> 1 \therefore p_F \approx p_N$.

Имеем статистику $F = \dfrac{s_1^2}{s_2^2}; \left\{\dfrac{\nu_1}{\nu_2}\right\}$. Тогда порядок проверки гипотезы H_0:

1) При заданном α найдем по таблице Фишера—Снедекора критическое значение $F_\alpha (P\{F > F_\alpha(\alpha, \nu_1, \nu_2)\} = \alpha)$.
2) Если $\widehat{F} = \dfrac{s_1^2}{s_2^2} > F_\alpha$, то H_0 отклоняется; если $\widehat{F} = \dfrac{s_1^2}{s_2^2} \le F_\alpha$, то H_0 не отклоняется.

Альтернативные гипотезы: если $H_1: \sigma_1^2 \ne \sigma_2^2$, то используется двухсторонний критерий, если $H_1: \sigma_1^2 > \sigma_2^2$, то используется односторонний критерий.

З а м е ч а н и е 1. Сравнение дисперсий производилось без предположений о средних.

З а м е ч а н и е 2. Для сравнения (равенства) дисперсий более двух совокупностей используются критерии χ^2-Пирсона, Бартлетта, Кохрана, Хартли и т. д., где вводятся соответствующие статистики для дисперсий. Их называют также критериями *однородности*.

17.4. Сравнение двух зависимых выборок (парные сравнения)

Пусть имеются два ряда наблюдений $\begin{matrix} X: X_1, X_2, \ldots, X_n \\ Y: Y_1, Y_2, \ldots, Y_n \end{matrix}$ и пусть наблюдения связаны попарно (например, производительность труда до и после внедрения новой технологии).

Имеем: n связанных пар (X_i, Y_i).

Задача: установить, когда различие между X_i и Y_i можно отнести за счет случайных отклонений, а когда за счет влияния изучаемого фактора.

Введем $d_i \equiv X_i - Y_i \therefore \hat{\bar{d}}_n = \dfrac{\sum_{i=1}^{n} d_i}{n}$. Логично проверить гипотезу

$H_0: \bar{d} = 0$ против $H_1: \bar{d} \gtrless 0$ или $H_1: \bar{d} \neq 0$.

Критерий проверки: статистика

$$t = \frac{\hat{\bar{d}}_n}{s\left(\hat{\bar{d}}_n\right)}, \text{ где } s\left(\hat{\bar{d}}_n\right) = \sqrt{\frac{\sum\left(d_i - \hat{\bar{d}}_n\right)^2}{n-1}}.$$

Пусть $(X_i - Y_i) \in N(0, \sigma)$, тогда t имеет распределение Стьюдента с $\nu = n - 1$.

Далее проверка аналогична п. 17.2.

18. Элементы непараметрического статистического вывода. Критерии согласия

18.1. Непараметрические сравнения двух выборок

Непараметрические методы не требуют знания закона генерального распределения изучаемого признака.

1. Сравнение двух независимых выборок.

$$\left.\begin{aligned} &X: X_1,\ X_2,\ldots,\ X_i,\ldots,\ X_{n_1} \\ &Y: Y_1,\ Y_2,\ldots,\ Y_j,\ldots,\ Y_{n_2} \end{aligned}\right\}\text{выборки.}$$

H_0: выборки из одной совокупности.

Для проверки гипотезы применяются критерии:

а) Критерии положения. H_0 *отклоняется* с $\alpha = 0{,}05$, если при $n = n_1 = n_2$ будет наблюдаться одно из: 1) не менее 5 наибольших или наименьших значений $\{X_i\}$, $\{X_j'\}$ при $n \le 25$ или 6 значений при $n > 25$ содержит одна и та же выборка; 2) не менее 7 значений одной выборки с большим размахом лежат вне размаха другой выборки.

б) Критерий медианы. Для объединенного вариационного ряда ищется M_e, определяется число наблюдений в каждой выборке, меньших и больших M_e (m_{11}, m_{12} — для первой и m_{21}, m_{22} — для второй), определяются теоретические частоты m'_{ij} из условия совпадения медиан выборок: $m'_{11} = m'_{12} = 0{,}5n_1$ и $m'_{21} = m'_{22} = 0{,}5n_2$. Далее оценивается расхождение теоретической и опытной частот с помощью χ^2 (см. п. 16.1 — сравнение долей двух малых выборок).

в) Ранговые критерии. Критерий У*илкоксона*. Измеренные значения признака заменяются на их порядковые номера — ранги.

Критерий $\boxed{u = \dfrac{\hat{\hat{R}} - E\left(\hat{\hat{R}}\right)}{\sigma\left(\hat{\hat{R}}\right)}}$,

где $E\left(\hat{\hat{R}}\right) = \dfrac{n_1 + n_2 + 1}{2}$, $\sigma\left(\hat{\hat{R}}\right) = \sqrt{\dfrac{(n_1 + n_2 + 1)n_2}{12n_1}}$.

$\hat{\hat{R}}$ — среднее значение ранга для наблюдений одной выборки в объединенном ранжированном ряду с рангом R каждого значения. Здесь ранги имеют значения 1, 2, ... , $(n_1 + n_2)$.

При связанных рангах (одинаковые значения признака) назначается средний ранг. Это влияет только на

$$\sigma\left(\hat{\hat{R}}\right)=\sqrt{\frac{\left(n^2-1\right)n-\sum T_i}{12n_1n}\cdot\frac{n_2}{n-1}},$$

где $n = n_1 + n_2$; $T_i = (t_i - 1)t_i(t_i + 1)$; t_i — число равных рангов в одной «связке».

При n_1, $n_2 \geq 20$ $u \tilde{\in} N(0, 1)$. Задается α, определяется z_α и проверяется $\hat{u} \overset{?}{\in} \left(-z_\alpha, z_\alpha\right)$.

2. Ранговый критерий парных сопоставлений.

Попарно связанные наблюдения, как в п. 17.4.

Разности $|d_i|$ располагают в неубывающем порядке. $d_i = 0$ не учитываются; $\min|d_i| \leftrightarrow 1$. Любому рангу R приписывают соответствующий знак d_i. Проверка гипотезы по критерию Уилкоксона

$u = \dfrac{\hat{R}-E(R)}{\sigma(R)}$, где $\hat{R}$ — сумма положительных рангов.

$$E(R)=\frac{n(n+1)}{4};\ \sigma^2(R)=\frac{n(n+1)(2n+1)}{24};$$

$$\sigma^2(R)=\frac{n(n+1)(2n+1)}{24}-\frac{\sum T_i}{48}$$

При $n > 25$ $u \tilde{\in} N(0, 1)$. Далее по известной схеме: $\alpha \to z_\alpha \to \hat{u} \overset{?}{\in} \left(-z_\alpha,\ z_\alpha\right)$.

18.2. Критерии согласия

Предназначены для проверки гипотезы о согласованности выборочного распределения с теоретическим (генеральным). Особое значение имеет проверка гипотезы о нормальном распределении $H_0:\ F_n^*(x) = F(x)$.

1. Критерий χ^2 Пирсона.

$\boxed{\chi^2 = \sum_{i=1}^{l} \frac{\left(\widehat{m}_i - m_i\right)^2}{m_i}}$, где ℓ — число классов (групп), на которое разбито опытное распределение; $\widehat{m}_i$ — частота вариант в i-й группе; m_i — теоретическая частота, рассчитанная по гипотетическому теоретическому распределению.

[$m_i = np_i$, для нормального распределения $m_i = n\Delta\Phi$].

Здесь число степеней свободы $\nu = \ell - k - 1$, где k — число параметров.

Схема проверки:

$$\alpha \xrightarrow{\text{по таблице}} \chi^2_\alpha \to \widehat{\chi}^2 \overset{?}{\lesseqgtr} \chi^2_\alpha \to H_0 \text{ при } \widehat{\chi}^2 < \chi^2_\alpha.$$

Замечание 1. Условие для применимости критерия: 1) выборка случайная; 2) $n \geq 50$, $m_i \geq 5 \div 10$.

2. Критерий Романовского (при фиксированном $\alpha = 0{,}027$).

$\boxed{r = \frac{\chi^2 - \nu}{\sqrt{2\nu}}}$, где $r \cong N(0,\, 1)$. Тогда $\left|\hat{r}\right| \leq 3 \Rightarrow H_0$ не отвергается.

3. Критерий Колмогорова.

$\boxed{D = \sqrt{n}\left(\max_i \left|F_n^*(x_i) - F(x_i)\right|\right)}$. При $n < 35$ при проверке гипотезы используется критическое значение D (по таблицам), при $n \geq 35$ используется предельное распределение Колмогорова. Схема проверки прежняя.

4. Критерий w^2 Смирнова-Крамера-Мизеса.

$$\boxed{W = n\omega^2 = n\sum_{i=1}^{l}\left(F_n^*(x_i) - F(x_i)\right)^2} = n\int_{-\infty}^{\infty}\left(F_n^*(x) - F(x)\right)dF(x).$$

Функция nw^2 табулирована для различных α. Схема проверки та же.

19. Элементы планирования эксперимента и дисперсионного анализа. Введение в факторный анализ

19.1. Модели эксперимента

Цель планирования эксперимента — получение более полного объема информации по сравнению с обычными методами при тех же затратах.

Данные контрольного эксперимента подвергаются специальному статистическому анализу — *дисперсионному анализу* (ДА).

Его суть заключается: 1) в расчленении общей дисперсии признака на компоненты согласно влиянию конкретных факторов и 2) в проверке гипотез о значимости их влияния.

Основное допущение: значение результата эксперимента можно представить в виде суммы ряда компонент.

Случай А. Исследуется влияние одного фактора, тогда модель структуры результата эксперимента:

$$x_{ij} = \bar{x} + \alpha_j + \varepsilon_{ij},$$

где x_{ij} — i-е значение признака, полученное на j-м уровне фактора (уровень — мера, количество); $\bar{x}$ — общая средняя; α_j — эффект фактора на j-м уровне; e_{ij} — случайная компонента, вызванная влиянием всех других факторов.

Случай В. Модель эксперимента предусматривает влияние нескольких факторов и их взаимодействие. Для двух факторов (A и B) структура результативного признака имеет модель

$$x_{ijk} = \bar{x} + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk},$$

где x_{ijk} — результат эксперимента в k-м наблюдении на i-м уровне A и j-м уровне B; α_i — эффект фактора A; β_j — эффект фактора B; γ_{ij} — совместный эффект, т. е. эффект взаимовлияния факторов A и B; ε_{ijk} — случайная компонента.

Замечание 1. Следовательно, при дисперсионном анализе совокупность разбивается на группы, отличающиеся по уровню факторов.

Замечание 2. Вводится допущение: 1) признаки приблизительно нормальны; 2) дисперсии в группах одинаковы.

19.2. Однофакторный анализ при полностью случайном плане эксперимента

Полностью случайный план эксперимента: изучается только влияние одного фактора. Пусть результаты наблюдения (N) разбиты на p групп, различающихся между собой по уровню фактора. Число наблюдений в j-й группе — n_j.

Таблица 1

Номер выборки	Наблюдаемые значения признака	Объем выборки	Сумма	Групповая средняя
1	$x_{11}, x_{21}, \ldots, x_{i1}, \ldots, x_{n_1 1}$	n_1	$\sum_i x_{i1} = T_1$	$\bar{x}_1 = \frac{1}{n_1}\sum x_{i1} = T_1/n_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	$x_{1j}, x_{2j}, \ldots, x_{ij}, \ldots, x_{n_j j}$	n_j	$\sum_i x_{ij} = T_j$	$\bar{x}_j = T_j/n_j$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
p	$x_{1p}, x_{2p}, \ldots, x_{ip}, \ldots, x_{n_p p}$	n_p	$\sum_i x_{ip} = T_p$	$\bar{x}_p = T_p/n_p$
	Итого	$N = \sum_i n_j$	$\sum_i \sum_j x_{ij} = G$	$\bar{x} = \frac{1}{N}\sum_i \sum_j x_{ij} = \frac{1}{p}\sum_1^p \bar{x}_j$

Поскольку имеем случай A, то требуется найти *межгрупповую* дисперсию и *внутригрупповую*. Первая обусловлена *влиянием изучаемого фактора*, вторая — *случайностью*.

1) Пусть имеем общую сумму квадратов отклонений:

$$\underbrace{\sum_i \sum_j (x_{ij} - \bar{x})^2}_{Q_0} = \underbrace{\sum_i \sum_j (x_{ij} - \bar{x}_j)^2}_{Q_1} + \underbrace{\sum_j n_j (\bar{x}_j - \bar{x})^2}_{Q_2}$$

2) Для получения несмещенных оценок дисперсий необходимо каждую из сумм разделить на число степеней свободы v_0, v_1, v_2 соответственно.

Очевидно: $v_0 = N - 1$; $v_1 = N - p$; $v_2 = p - 1$.

Тогда несмещенные оценки $s_0^2 = \dfrac{Q_0}{N-1}$; $s_1^2 = \dfrac{Q_1}{N-p}$; $s_2^2 = \dfrac{Q_2}{p-1}$,

где s_1^2 — характеризует рассеяние внутри группы, s_2^2 — рассеяние групповых средних.

3) Задачу различия дисперсий можно рассматривать как эквивалентную задаче проверки существенности различия между выборками. В самом деле, если влияние фактора отсутствует, то s_1^2 и s_2^2, можно рассматривать как независимые оценки дисперсии совокупности σ^2, а если фактор оказывает заметное влияние, то s_2^2 / s_1^2 больше критического предела, а следовательно, выборки взяты из разных совокупностей.

$H_0 : \sigma_1^2 = \sigma_2^2$. По выборочным данным вычисляются s_1^2 и s_2^2; используется статистика

$$F = \frac{s_2^2}{s_1^2};\ \left\{\begin{matrix} p-1 \\ N-p \end{matrix}\right\} \to \text{задается } \alpha \xrightarrow{\text{по таблице}} F_\alpha \to \widehat{F} \le F_\alpha \Rightarrow H_0$$

не отклоняется.

Таблица 2

Характер вариации	**Сумма квадратов**	**Число степеней свободы**	**Оценка дисперсии**
Систематическая (межгрупповая)	$Q_2 = \sum_j n_j \left(\bar{x}_j - \bar{x}\right)^2$	$p-1$	$s_2^2 = \dfrac{Q_2}{p-1}$
Остаточная (внутригрупповая)	$Q_1 = \sum_i \sum_j \left(x_{ij} - \bar{x}_j\right)^2$	$N-p$	$s_1^2 = \dfrac{Q_1}{N-p}$
Итого	$Q_0 = \sum_i \sum_j \left(x_{ij} - \bar{x}\right)^2$	$N-1$	—

С учетом обозначений табл. 1:

$$Q_0 = \sum_i \sum_j x_{ij}^2 - \frac{G^2}{N};\ Q_2 = \sum_j T_j^2 - \frac{G^2}{N} \text{ и } Q_1 = Q_0 - Q_2.$$

19.3. Однофакторный анализ при группировке по случайным блокам

Суть метода планирования эксперимента — «метода случайных блоков» — заключается в предварительном разбиении наблюдаемого пространства на «блоки» с примерно одинаковыми элементами внутри каждого. Затем каждый блок разбивается на число групп, совпадающих с количеством уровней фактора. Число единиц наблюдения каждого уровня фактора должно быть одинаковым, т. е. $n_1 = n_2 = \ldots = n_p = n$. Распределение уровней фактора по группам случайно.

Модель экспериментального результата: $x_{ij} = \bar{x} + \alpha_i + \beta_j + \varepsilon_{ij}$, где α_i — эффект блоков; β_j — эффект уровня фактора.

При методе «случайных блоков» уменьшается разброс наблюдений.

Таблица 3

Уровень фактора	**Результаты наблюдения по блокам**						**Сумма по строкам**	**Средняя по уровню фактора**
	1	2	…	i	…	n		
1	x_{11}	x_{21}	…	x_{i1}		x_{n1}	T_1	$\bar{x}_1$
2	x_{12}	x_{22}	…	x_{i2}		x_{n2}	T_2	$\bar{x}_2$
…	…	…	…	…		…	…	…
j	x_{1j}	x_{2j}	…	x_{ij}		x_{nj}	T_j	$\bar{x}_j$
…	…	…	…	…		…	…	
p	x_{1p}	x_{2p}	…	x_{ip}		x_{np}	T_p	$\bar{x}_p$
Сумма по вертикали	B_1	B_2	…	B_i		B_n	G	$\bar{x}$
Средняя по блокам	$\bar{B}_1$	$\bar{B}_2$	…	$\bar{B}_i$		$\bar{B}_n$	—	—

Сумма квадратов отклонений здесь должна быть разбита на 3 составляющие.

$$\sum_{i,j}\left(x_{ij} - \bar{x}\right)^2 = \sum_{i,j}\left(x_{ij} - \bar{B}_i + \bar{B}_i - \bar{x}_j + \bar{x}_j - \bar{x} + \bar{x} - \bar{x}\right)^2 =$$

$$= \sum_{i,j}\left[\left(x_{ij} - \bar{B}_i - \bar{x}_j + \bar{x}\right) + \left(\bar{B}_i - \bar{x}\right) + \left(\bar{x}_j - \bar{x}\right)\right]^2 =$$

$$= \sum_{i,j}\left(x_{ij} - \bar{B}_i - \bar{x}_j + \bar{x}\right)^2 + \sum_{i,j}\left(\bar{B}_i - \bar{x}\right)^2 + \sum_{i,j}\left(\bar{x}_j - \bar{x}\right)^2 =$$

$$= \underbrace{\sum_{i,j}\left(x_{ij} - \bar{B}_i - \bar{x}_j + \bar{x}\right)^2}_{Q_1 \text{ — остат.вариация}} + \underbrace{p\sum_{i}\left(\bar{B}_i - \bar{x}\right)^2}_{Q_2 \text{ — вар. между блоками}} + \underbrace{n\sum_{j}\left(\bar{x}_j - \bar{x}\right)^2}_{\substack{Q_3 \text{ — межгрупп. вариация} \\ \text{межуровневая}}}$$

Вычисления: Если $Q_0 \equiv \sum_{i,j}\left(x_{ij} - \bar{x}\right)^2 \Rightarrow Q_1 = Q_0 - \left(Q_2 + Q_3\right)$,

где $Q_0 = \sum_{i,j} x_{ij}^2 - \dfrac{G^2}{N}$; $Q_2 = \sum_i \dfrac{B_i^2}{p} - \dfrac{G^2}{N}$; $Q_3 = \sum_j \dfrac{T_j^2}{n} - \dfrac{G^2}{N}$.

Число степеней свободы для: $Q_0 \to N-1$; $Q_2 \to n-1$; $Q_3 \to p-1$; $Q_1 \to \nu_1 = \nu_0 - \left(\nu_2 + \nu_3\right) = (n-1)(p-1)$.

Далее используется критерий $F = \dfrac{s_2^2}{s_1^2}$; $\left\{\dfrac{p-1}{(n-1)(p-1)}\right\}$; проверяется $H_0:\ \sigma_1^2 = \sigma_3^2$; $\alpha \to F_\alpha \to \widehat{F} \le F_\alpha \to H_0$ не отклоняется.

Таблица 4

Характер вариации	**Сумма квадратов**	**Число степеней свободы**	**Оценка дисперсии**
Систематическая (межгрупповая)	$Q_3 = n\sum_j\left(\bar{x}_j - \bar{x}\right)^2$	$p-1$	$s_3^2 = \dfrac{Q_3}{p-1}$
Между блоками	$Q_2 = \sum_i p\left(\bar{B}_i - \bar{x}\right)^2$	$n-1$	$s_2^2 = \dfrac{Q_2}{n-1}$
Остаточная (внутригрупповая)	$Q_1 = Q_0 - \left(Q_2 + Q_3\right)$	$(n-1)(p-1)$	$s_1^2 = \dfrac{Q_1}{(n-1)(p-1)}$
Итого	$Q_0 = \sum_i\sum_j\left(x_{ij} - \bar{x}\right)^2$	$N-1$	—

19.4. Двухфакторный анализ при полностью случайном плане эксперимента

Пусть выявляется роль двух факторов A и B и их взаимодействия на некоторый результативный признак.

Модель эксперимента: $x_{ijk} = \overline{x} + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$.

Опыт производится при фиксированных A и B. Опыт повторяется n раз для каждого сочетания факторов. Следовательно, получаем n значений признака. Данные эксперимента сводятся в таблицу.

Таблица 5

Уровень фактора B	**Уровень фактора A**				**Сумма**
	A_1	A_2	...	A_p	
B_1	$x_{111}, x_{112}, \dots, x_{11n}$	$x_{211}, x_{212}, \dots, x_{21n}$	...	$x_{p11}, x_{p12}, \dots, x_{p1n}$	$\sum\limits_{i,k} x_{i1k}$
B_2	$x_{121}, x_{122}, \dots, x_{12n}$	$x_{221}, x_{222}, \dots, x_{22n}$		$x_{p21}, x_{p22}, \dots, x_{p2n}$	$\sum\limits_{i,k} x_{i2k}$
...	...	...	...	...	...
B_q	$x_{1q1}, x_{1q2}, \dots, x_{1qn}$	$x_{2q1}, x_{2q2}, \dots, x_{2qn}$	...	$x_{pq1}, x_{pq2}, \dots, x_{pqn}$	$\sum\limits_{i,k} x_{iqk}$
Сумма	$\sum\limits_{j,k} x_{1jk}$	$\sum\limits_{j,k} x_{2jk}$	...	$\sum\limits_{j,k} x_{pjk}$	$\sum\limits_{i,jk} x_{ijk}$

По данным таблицы определяются:

1) общая средняя $\overline{X} = \dfrac{\sum\limits_{ijk} X_{ijk}}{npq}$;

2) средняя по строкам $\overline{T}_j = \dfrac{\sum\limits_{i,k} x_{ijk}}{pn}$;

3) средняя по столбцам $\overline{T}_i = \dfrac{\sum\limits_{jk} x_{ijk}}{qn}$;

4) средняя для сочетания уровней факторов $\bar{X}_{ij} = \dfrac{\sum\limits_k x_{ijk}}{n}$.

По модели эксперимента имеем четыре составляющие суммы квадратов отклонений

$$\sum_{ijk}\left(x_{ijk}-\bar{X}\right)^2 = \underbrace{\sum_{ijk}\left(\bar{T}_i-\bar{X}\right)^2}_{\substack{\parallel \\ \text{Вл. ф}\,A \to Q_4}} + \underbrace{\sum_{ijk}\left(\bar{T}_j-\bar{X}\right)^2}_{\substack{\parallel \\ \text{Вл. ф}\,B \to Q_3}} + \underbrace{\sum_{ijk}\left(\bar{X}_{ij}-\bar{T}_i-\bar{T}_j+\bar{X}\right)^2}_{\substack{\parallel \\ \text{Взаим. } A \text{ и } B \to Q_2}} + \underbrace{\sum_{ijk}\left(x_{ijk}-\bar{X}_{ij}\right)^2}_{\substack{\parallel \\ \text{Остат. в.} \to Q_1}}$$

Таблица 6

Источник вариации	**Сумма квадратов**	**Число степеней**	**Оценка дисперсии свободы**
Фактор A	$Q_4 = nq\sum\limits_i\left(\bar{T}_i-\bar{X}\right)^2$	$p-1$	$s_4^2 = \dfrac{Q_4}{p-1}$
Фактор B	$Q_3 = np\sum\limits_j\left(\bar{T}_j-\bar{X}\right)^2$	$q-1$	$s_3^2 = \dfrac{Q_3}{q-1}$
Взаимодействие $A\times B$	$Q_2 = n\sum\limits_{ij}\left(X_{ij}-\bar{T}_i-\bar{T}_j+\bar{X}\right)^2$	$(p-1)(q-1)$	$s_2^2 = \dfrac{Q_2}{(p-1)(q-1)}$
Остаточная вариация	$Q_1 = \sum\limits_{ijk}\left(x_{ijk}-\bar{X}_{ij}\right)^2$	$N-pq$	$s_1^2 = \dfrac{Q_1}{N-pq}$
Итого	$Q_0 = \sum\limits_{ijk}\left(x_{ijk}-\bar{X}\right)^2$	$N-1$	—

$N = npq$. Число степеней свободы соответственно:
1) между столбцами $\nu_4 = p-1$;
2) между строками $\nu_3 = q-1$;
3) для взаимодействия $\nu_2 = (p-1)(q-1)$;
4) внутри ячеек $(n-1)pq = N-pq$.

Далее осуществляется проверка гипотез по известной схеме:

1) $\hat{F}_A = \dfrac{s_4^2}{s_1^2}$; $\left\{\dfrac{p-1}{N-pq}\right\}$;

2) $\widehat{F}_B = \frac{s_3^2}{s_1^2}$; $\left\{\frac{q-1}{N-pq}\right\}$;

3) $\widehat{F}_{AB} = \frac{s_2^2}{s_1^2}$; $\left\{\frac{(p-1)(q-1)}{N-pq}\right\}$.

Обычно используют: $Q_4 = \left(\sum_i^p \frac{T_i^2}{nq}\right) - \frac{G^2}{N}$; $Q_3 = \left(\sum_j^q \frac{T_j^2}{np}\right) - \frac{G^2}{N}$;

$$Q_2 = \sum_{ij} \frac{T_{ij}^2}{np} - \sum_i^p \frac{T_i^2}{nq} - \sum_j^q \frac{T_j^2}{np} + \frac{G^2}{N};\ Q_1 = \sum_{ij}\left(\sum_k^n x_{ijk}^2 - \frac{T_{ij}^2}{n}\right)$$

$$G = \sum_{ijk} x_{ijk};\ T_{ij} = \sum_k^n x_{ijk},\ T_i \text{ и } T_j \text{ — суммы по } i \text{ и } j.$$

20. Основы теории корреляции и регрессии

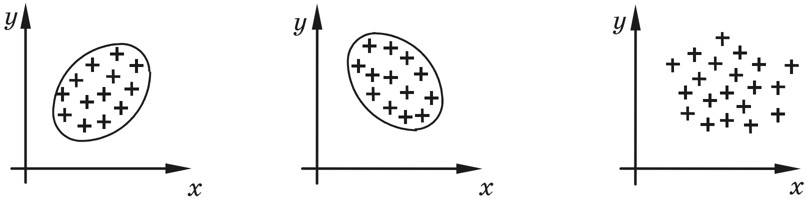
20.1. Основные понятия и определения

Статистический анализ зависимости не выявляет существо причинных связей между явлениями. Причинный анализ является результатом качественного (содержательного) изучения связей, которое предшествует или сопровождает статистический анализ.

Определение 1. Связь условной средней одной величины от соответствующих значений другой величины называется ***корреляционной связью***, а уравнение связи $\overline{y}_k = f(x_k)$ — ***уравнением регрессии*** на x.

Иначе: частный случай стохастической зависимости, при которой изменение одной случайной величины X приводит к изменению условной средней другой случайной величины Y, называется ***корреляционной*** зависимостью. Связь или корреляция двух переменных называется ***парной***. Если с ростом величины X величина Y в среднем растет, то корреляция **положительная**, если Y убывает, то корреляция **отрицательная**.

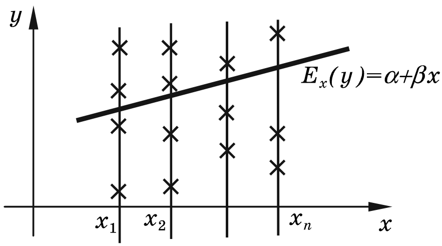
Определение 2. ***Корреляционным полем*** называется диаграмма, изображающая совокупность значений двух признаков.



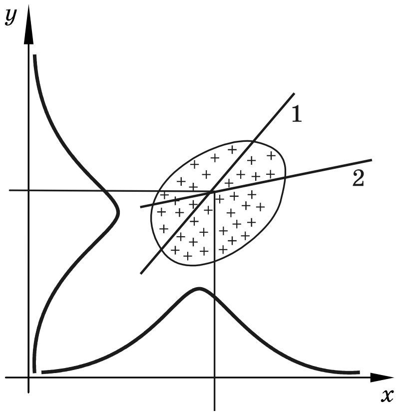
Замечание 1. Совокупность двух признаков может представляться через числовые данные наблюдений в виде таблицы, которую называют ***корреляционной матрицей***.

При изучении зависимости используют *два подхода (модели)*.

1. ***Регрессионная*** модель. Экспериментатор задает значение величины X и наблюдает Y. Тогда каждому X соответствует распределение Y с дисперсией σ^2. Наблюдения рассматриваются как выборочные значения. $E(y)=\alpha+\beta x$ — ***уравнение регрессии***.



2. ***Корреляционная*** модель. Здесь X и Y есть выборки из двумерного нормального распределения, X и Y не фиксированные.



Линии регрессии:

1. $E_x(y)=\alpha+\beta x$.
2. $E_y(y)=\alpha_1+\beta_1 y$.

20.2. Уравнение парной регрессии

Построение уравнения регрессии связано с решением двух задач:

1. Выбора независимых переменных и определения вида уравнения регрессии (этап *спецификации*).
2. Оценивания параметров (коэффициентов) уравнения.

Первая задача решается на основе качественного анализа изучаемой связи с принятием вида уравнений (линейного, гиперболического, параболического, логического типа). Можно использовать графики.

Вторая задача связана с последующей проверкой оценок параметров.

(1) *Линейное уравнение парной регрессии* $y=\alpha+\beta x+\varepsilon$, где α, β — неизвестные коэффициенты регрессии, ε — случайная переменная (возмущение), $\overline{y}=\alpha+\beta x$ — систематическая часть.

Пусть $\varepsilon\in N\left(0,\ \sigma_\varepsilon^2=const\right)$ и $\mathrm{E}\varepsilon_i\varepsilon_j=0$ при $i\neq j$.

Для любого наблюдения: $y_i=\alpha+\beta x_i+\varepsilon_i$. Задаются X_i и наблюдаются Y_i, строятся статистические оценки для α и β из системы уравнений относительно α и β. Следовательно $\hat{\overline{y}}=\mathrm{a}+bx$ есть оценка модели $y=\alpha+\beta x+\varepsilon$.

(2) *Метод наименьших квадратов (МНК) для парной регрессии* (минимум суммы квадратов отклонений наблюдений от линии регрессии).

$e_i=Y_i-\hat{y}_i=Y_i-\left(\mathrm{a}+bX_i\right)$, где Y_i — фактические наблюдения, $\hat{y}_i$ — соответствующие им расчетные значения. Сумма квадратов отклонений $I\equiv\sum\limits_i e_i^2=\sum\limits_i\left(Y_i-\mathrm{a}-bX_i\right)^2$. Минимум достигается при $\dfrac{\partial I}{\partial \mathrm{a}}=0$ и $\dfrac{\partial I}{\partial b}=0$, т. е.

$$\left.\begin{array}{l}\dfrac{\partial I}{\partial \mathrm{a}}=-2\sum\limits_i\left(Y_i-\mathrm{a}-bX_i\right)=0\\ \dfrac{\partial I}{\partial b}=-2\sum\limits_i\left(Y_i-\mathrm{a}-bX_i\right)X_i=0\end{array}\right\}\to\left.\begin{array}{c}\sum\limits_i Y_i=n\mathrm{a}+b\sum\limits_i X_i\\ \sum\limits_i X_iY_i=\mathrm{a}\sum\limits_i X_i+b\sum\limits_i X_i^2\end{array}\right\}\Rightarrow$$

$$\Rightarrow\begin{array}{l}\hat{\overline{Y}}_n=\mathrm{a}+b\hat{\overline{X}}_n\\ \mathrm{a}=\hat{\overline{Y}}_n-b\hat{\overline{X}}_n\end{array},\quad b=\frac{\sum X_iY_i-n\hat{\overline{X}}_n\hat{\overline{Y}}_n}{\sum\limits_i X_i^2-n\hat{\overline{X}}_n^2}.$$

(3) *Свойства оценок МНК.*

а) Несмещенность $E\mathrm{a}=\alpha$ и $Eb=\beta$.

б) Состоятельность $\lim\limits_{n\to\infty}\sigma_a^2=0$ и $\lim\limits_{n\to\infty}\sigma_b^2=0$.

в) Эффективность. Имеет место, если в п. **(1)** $\sigma_\varepsilon^2=const$ и $\mathrm{E}\left(\varepsilon_i\varepsilon_j\right)=0$.

(4) *Ошибки коэффициентов регрессии.*

Сумма квадратов отклонений от линии регрессии $I=\sum e_i^2$. Несмещенная оценка дисперсии относительно регрессии есть $s^2=\dfrac{I}{n-2}$, т. к. две степени свободы «теряются» при определении a и b. Т. е. s^2 есть выборочная оценка дисперсии возмущений ε_i.

Можно найти:

$$D(b)=s_b^2=\frac{s^2}{\sum\left(X_i-\hat{\bar{X}}_n\right)^2},\quad D(\mathrm{a})=s_a^2=\frac{\sum X_i^2}{n\sum\left(X_i-\hat{\bar{X}}_n\right)^2}s^2.$$

Далее можно проверить значимость a и b по схеме проверки статистических гипотез. Обычно проверяют значимость b H_0: $\beta=0$ по статистике $t=\dfrac{b}{s_b}$ по Стьюденту при $\nu=n-2$. Если $\hat{t}>t_\alpha$, то H_0 — отклоняется.

(5) *Доверительные интервалы линии регрессии.*

Для случая $\hat{y}_i=\hat{\bar{Y}}+b\left(X_i-\hat{\bar{X}}_n\right)$ имеем $s_{\hat{y}}^2=\dfrac{s^2}{n}+\dfrac{s^2\left(x_p-\hat{\bar{X}}_n\right)^2}{\sum\left(X_i-\hat{\bar{X}}_n\right)^2}$, где

x_p — значение x, для которого определяется $\hat{y}$.

$\min\limits_{x_p}s_{\hat{y}}^2=\dfrac{s^2}{n}$ при $x_p=\hat{\bar{X}}$.

Тогда доверительные границы есть $\left(\hat{y}_i\pm t_\alpha s_{\hat{y}}^2\right)$.

20.3. Коэффициент корреляции

С помощью коэффициента корреляции двух случайных величин можно измерять степень линейной зависимости признаков в генеральной совокупности.

О п р е д е л е н и е 3. Для выборочных данных эмпирическая мера связи X и Y определяется через *выборочный **коэффициент корреляции***, определяемый по правилу

$$\hat{r} = \frac{c\hat{o}v(X,Y)}{s_{\bar{x}} s_{\bar{y}}} = \frac{\sum_i \left(X_i - \hat{\bar{X}}_n\right)\left(Y_i - \hat{\bar{Y}}_n\right)}{\sqrt{\sum_i \left(X_i - \hat{\bar{X}}_n\right)^2 \cdot \sum_i \left(Y_i - \hat{\bar{Y}}_n\right)^2}} =$$

$$= \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{\left[n\sum X_i^2 - \left(\sum X_i\right)^2\right]\left[n\sum Y_i^2 - \left(\sum Y_i\right)^2\right]}}.$$

З а м е ч а н и е 2. Очевидно $-1 \le \hat{r} \le 1 \Rightarrow \hat{r} = 1$ (линейная связь с положительным углом наклона), $\hat{r} = -1$ (обратная линейная связь), $\hat{r} = 0$ (отсутствует линейная связь).

При нормально распределенных X и Y $\hat{r}$ есть мера *линейной* согласованности X и Y.

З а м е ч а н и е 3. При модели с фиксированными X $\hat{r}$ есть мера близости эмпирических точек к линии регрессии.

Связь между параметром парной регрессии b и *коэффициентом корреляции* $\hat{r}$: $b = \hat{r}\dfrac{s_{\bar{Y}}}{s_{\bar{x}}}$, где $s_{\bar{Y}}$, $s_{\bar{x}}$ — среднеквадратические отклонения.

Можно показать, что $\hat{r}$ определяется через b_{yx} и b_{xy} как

$$\hat{r} = \sqrt{b_{yx} b_{xy}}.$$

Если зависимость между признаками функциональная, то

$$b_{yx} = \frac{1}{b_{xy}} \Rightarrow \hat{r} = 1.$$

Проверка значимости $\hat{r}$. Проверяется гипотеза H_0: $r = 0$.

Вводится статистика Стьюдента $t = \dfrac{\hat{r}\sqrt{n-2}}{\sqrt{1-\hat{r}^2}} \xrightarrow{\text{по формуле}} \hat{t}$, сопоставляется с табличной t_α при $n-2$ степенях свободы и заданной α.

Замечание 4. Так как величина t полностью определяется числом наблюдений n и величиной $\hat{r}$, то можно найти $\hat{r}_{\min}$, при котором H_0: $r = 0$ может быть отклонена с заданной вероятностью.

Пример: $n = 42$; $\hat{r} = 0{,}65 \Rightarrow \hat{t} = 5{,}27$; при $\alpha = 0{,}05$ и $(n-2) = 40$; $t_\alpha = 2{,}02$ и $r \neq 0$.

При $\hat{r}_{\min} = 0{,}30$ имеем отклонение гипотезы H_0.

Определение 4. ***Коэффициент детерминации*** есть величина, равная квадрату коэффициента корреляции. Это есть мера качества подбора линии регрессии.

Если зависимую переменную представить в виде $Y_i = \bar{Y} + k_i + e_i$, где $k_i = b\left(X_i - \bar{X}\right)$ — систематическая составляющая от уравнения регрессии, $e_i = \left(Y_i - \bar{Y} - k_i\right)$ — случайная составляющая, $\hat{y}_i \equiv bX_i$, то можно записать для суммы квадратов отклонений:

$$\sum\left(Y_i - \bar{Y}\right)^2 = \sum k_i^2 + \sum e_i^2,$$

и тогда с учетом значений k и b

$$\frac{\sum k_i^2}{\sum\left(Y_i - \bar{Y}\right)^2} = \frac{\sum\left(\hat{y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} = \hat{r}^2 \Rightarrow \hat{r}^2 = \frac{s_{\hat{y}}^2}{s_{\bar{Y}}^2} = 1 - \frac{s^2}{s_{\bar{Y}}^2},$$

где $s_{\hat{y}}^2$ — дисперсия линии регрессии относительно средней;
s^2 — дисперсия остаточных членов относительно линии регрессии;
$s_{\bar{Y}}^2$ — общая дисперсия.

Величину $\hat{r} = \sqrt{1 - \dfrac{s^2}{s_{\bar{Y}}^2}}$ называют индексом корреляции.

20.4. Коэффициент ранговой корреляции

Иногда признаки не имеют абсолютной шкалы измерения (профессии, предпочтения, качественные особенности и т. д.) или практически их нельзя измерить. Тогда данные упорядочиваются по некоторому правилу (ранжируются). Ранг 1 приписывается наиболее важному объекту, ранг 2 — следующему и т. д. Если объекты ранжированы по двум признакам, то можно найти взаимосвязь.

Определение 5. ***Коэффициентом ранговой корреляции*** Спирмэна называют величину, определяемую по правилу

$$\hat{r}_S \equiv 1 - \frac{6\sum d_i^2}{n\left(n^2-1\right)},$$

где d_i — разность значений рангов, расположенных в двух рядах у одного и того же объекта, n — число данных (число рангов).

$$\overset{\text{полная обратная связь}}{-1} \le \hat{r}_S \le \overset{\text{полная прямая связь}}{+1}.$$

Замечание 5. Коэффициент $\hat{r}_S$ есть частный случай коэффициента парной корреляции. Коэффициент $\hat{r}_S$ можно получить из $\hat{r}$ заменой x и y на соответствующие ранги.

Пример 1.

Студент	А	Б	В	Г	Д	Е	Ж	З	И	К
Вступительный экзамен	2	5	6	1	4	10	7	8	3	9
Экзаменационная сессия	3	6	4	1	2	7	8	10	5	9
d_i	–1	–1	2	0	2	3	–1	–2	–2	0
d_i^2	1	1	4	0	4	9	1	4	4	0

$$\sum d_i^2 = 28,\ \underline{\hat{r}_S} = 1 - \frac{6\cdot 28}{10\left(10^2-1\right)} = \underline{0{,}83}.$$

Замечание 6. Если объекты связаны (неразличимы), то им приписывается одинаковый средний ранг.

Проверка значимости $\hat{r}_S$. H_0: $r_S = 0$. Полагают, что $\hat{r}_S$ стремится к нормальному закону. Среднеквадратическое отклонение определяется как $S_r = \dfrac{1}{\sqrt{n-1}}$. Пусть $\alpha = 0{,}05$, тогда:

а) H_0 — отклоняется, если $\hat{r}_S < \dfrac{-1{,}96}{\sqrt{n-1}}$ или $\hat{r}_S > \dfrac{1{,}96}{\sqrt{n-1}}$.

б) H_0 — не отклоняется, если $\hat{r}_S \geq \dfrac{-1{,}96}{\sqrt{n-1}}$ или $\hat{r}_S \leq \dfrac{1{,}96}{\sqrt{n-1}}$.

20.5. Коэффициент согласованности

Имеется несколько рядов рангов (работает группа экспертов).

Определение 6. ***Коэффициентом согласованности (конкордации)*** называют величину, определяемую по правилу

$$W = \frac{12\sum D_i^2}{m^2\left(n^3 - n\right)},$$

где n — число объектов; m — число экспертов (рядов рангов); D_i — отклонение суммы рангов объекта от средней их суммы для всех объектов ($0 \leq W \leq 1$).

20.6. Множественная линейная регрессия

Определение 7. Регрессия называется ***множественной***, если ее уравнение описывает связь какой-либо характеристики явления от нескольких факторов, т. е. зависимость одной величины от нескольких m.

В этом случае удобна матричная запись. Известная линейная модель тогда запишется в виде n-мерного вектора наблюдений:

$$\vec{y} = X\vec{\alpha} + \vec{\varepsilon},$$

где $\vec{y} = \left\{y_i\right\}_{i=1}^{n}$ — вектор значений зависимой переменной,

$X = \{x_{ij}\}$ — матрица значений независимых переменных ($n \times m$);

$\vec{\alpha} = \{\alpha_j\}_1^m$ — вектор неизвестных параметров;

$\vec{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$ — случайный вектор ошибок $\left(E\varepsilon_i = 0,\ D\varepsilon_i = \sigma^2,\ E\varepsilon_i\varepsilon_j = 0\right)$.

Уравнение множественной регрессии примет вид $\hat{\vec{y}} = X\vec{a}$, где $\vec{a} = \{a_j\}_{j=1}^m$ — вектор оценок параметров. Вектор отклонений оценки $\vec{e} = \vec{y} - X\vec{a}$.

Очевидно, сумма квадратов отклонений:

$$Q \equiv \vec{e}^T\vec{e} = (\vec{y} - X\vec{a})^T(\vec{y} - X\vec{a}) = \vec{y}^T\vec{y} - \vec{a}^t X^T\vec{y} - \vec{y}^T X\vec{a} + \vec{a}^T X^T X\vec{a}.$$

Подбор вектора оценок $\vec{a}$ осуществляется по методу наименьших квадратов, т. е. из требования $\frac{\partial Q}{\partial \vec{a}} = \left\{\frac{\partial Q}{\partial a_j}\right\}_{j=n}^m = \vec{0}$.

$$\frac{\partial Q}{\partial \vec{a}} = \frac{\partial}{\partial \vec{a}}\left(\vec{y}^T\vec{y} - 2\vec{a}^T X^T\vec{y} + \vec{a}^T X^T X\vec{a}\right) = -2X^T\vec{y} + 2X^T X\vec{a} = \vec{0}.$$

Отсюда $X^T\vec{y} = X^T X\vec{a} \Rightarrow \vec{a} = \left(X^T X\right)^{-1} X^T\vec{y}$.

Определение 8. Оценка вектора неизвестных коэффициентов линейной регрессии, определяемая по последней формуле, называется ***оценкой МНК***.

Пример 1. Пусть число независимых переменных $m = 2$, число измерений n.

Тогда $X = \begin{pmatrix} 1x_{11} & x_{12} \\ \vdots & \\ 1x_{n1} & x_{n2} \end{pmatrix}$; $\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$;

$$X^T X = \begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{pmatrix}; \; X^T \vec{y} = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \sum y_i x_{i2} \end{pmatrix}.$$

Замечание 7.

a) Оценка $\vec{a}$ оказывается несмещенной

и $E\vec{a} = \left(X^T X\right)^{-1} X^T \cdot XE\vec{\alpha} + E\vec{\varepsilon} = \vec{\alpha}.$

b) Сравнение коэффициентов регрессии, соизмеряющих влияние различных воздействующих на явление факторов, является корректным только при условии нормирования коэффициентов по формуле $\left\|a_j\right\| = a_j \dfrac{s_{x_j}}{s_y}$, где s_{x_j} — среднеквадратическое отклонение переменной x_j, s_y — среднеквадратическое отклонение для y.

Замечание 8. Коэффициент множественной корреляции можно определить по формуле:

$$R = \sqrt{1 - \frac{\sum e_i^2}{\sum\left(y_i - \bar{Y}\right)^2}} \text{ или } R = \sqrt{\frac{\vec{a}^T X^T \vec{y} - n\bar{Y}^2}{\vec{y}^T \vec{y} - n\bar{Y}^2}}.$$

20.7. Доверительные интервалы множественной регрессии

При оценке $\vec{a}$ вектора неизвестных коэффициентов по МНК относительно вектора ошибок предполагается $\varepsilon_i \in N(0,\sigma)$; $E\varepsilon_i\varepsilon_j = 0$ при $i \neq j$. Если столбцы матрицы X — линейно независимы, т. е. существует $\left(X^T X\right)^{-1}$, то оценки МНК — не смещены, состоятельны и эффективны.

Дисперсия оценки: $cov(\vec{a}) = \mathrm{E}\left[(\vec{a} - \vec{\alpha})(\vec{a} - \vec{\alpha})^T\right].$

Имеем

$$\vec{a}-\vec{\alpha}=\left(X^TX\right)^{-1}X^T\vec{y}-\vec{\alpha}=\left(X^TX\right)^{-1}X^T\left(X\vec{\alpha}+\vec{\varepsilon}\right)-\vec{\alpha}=\left(X^TX\right)^{-1}X^T\vec{\varepsilon}.$$

Тогда

$$cov\left(\vec{a}\right)=E\left[\left(X^TX\right)^{-1}X^T\vec{\varepsilon}\vec{\varepsilon}^TX\left(X^TX\right)^{-1}\right]=$$

$$=\underbrace{E\left(\vec{\varepsilon}\vec{\varepsilon}^T\right)}_{\substack{\| \\ \sigma^2\cdot I}}\cdot\left(X^TX\right)^{-1}=\sigma^2\cdot I\left(X^TX\right)^{-1}.$$

Таким образом $cov\left(\vec{a}\right)=\sigma^2\cdot\left(X^TX\right)^{-1}$.

Но σ^2 — неизвестная. Пусть $\vec{e}\equiv\vec{y}-\hat{\vec{y}}$, тогда оценка дисперсии

$$s^2=\frac{Q}{n-m-1}=\frac{\vec{e}^T\vec{e}}{n-m-1}=\frac{\sum e_i^2}{n-m-1}\Rightarrow cov\left(\vec{a}\right)=s^2\left(X^TX\right)^{-1}.$$

Доверительные интервалы регрессии. Уравнение регрессии $\hat{\vec{y}}=X\vec{a}$.

В нем $\vec{a}$ есть случайная величина, тогда и $\hat{\vec{y}}$ — случайная величина.

Тогда $D\left(\hat{\vec{y}}\right)=D\left(X\vec{a}\right)=Xcov\left(\vec{a}\right)X^T\cdot I=\sigma^2X\left(X^TX\right)^{-1}X^T\cdot I$

или $D\left(\hat{\vec{y}}\right)\approx s^2X\left(X^TX\right)^{-1}X^T\cdot I$.

Тогда доверительный интервал

$$\left(\hat{\vec{y}}-s\cdot t_\alpha\left[X\left(X^TX\right)^{-1}X^T\right]^{\frac{1}{2}}\cdot I,\ \hat{\vec{y}}+s\cdot t_\alpha\left[X\left(X^TX\right)^{-1}X^T\right]^{\frac{1}{2}}\cdot I\right),$$

где t_α — t-статистика по таблице при заданной α.

Прогнозное значение $\vec{y}:X\vec{a}$.

20.8. Нелинейная регрессия

Часто исследуемое явление наиболее адекватно описывают уравнения нелинейные как по переменным, так и по параметрам.

(1) Пусть необходимо рассчитать параметры уравнения

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_2^2 + e.$$

Введем переменные: $z_1 \equiv x_1$; $z_2 \equiv x_2$; $z_3 \equiv x_1^2$; $z_4 \equiv x_2^2$.

Тогда $y = a_0 + a_1 z_1 + a_2 z_2 + a_3 z_3 + a_4 z_4 + e.$

В векторной записи уравнение регрессии сводится к линейному $\vec{y} = Z\vec{a}$, и по аналогии с п. 20.6 имеем $\vec{a} = \left(Z^T Z\right)^{-1} Z^T \vec{y}.$

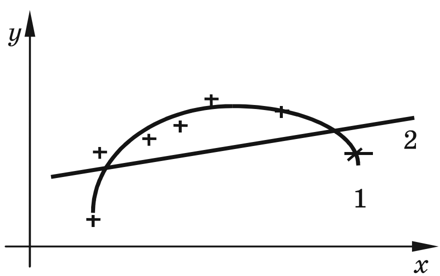
Ошибки параметров и доверительные интервалы ищутся как в п. 20.7.

Пример: Зарегистрированы данные, характеризующие интенсивность орошения x и урожайность y зерновой культуры.

y, ц/га	6	7	13	16	20	24	22	20
x, дм	0,9	1,0	1,8	2,4	4,0	5,8	7,6	8,5

Принимаем $y = a_0 + a_1 x + a_2 x^2 = a_0 + a_1 z_1 + a_2 z_2$.

Выполняя выкладки, имеем: $\hat{y} = 0{,}9127 + 7{,}7231x - 0{,}6638x^2$.



(2) Нелинейность модели по параметрам.

Например: $f(a_1, a_2) = a_1 x^{a_2}$, $f(a_1, a_2, a_3) = a_1 K^{a_2} L^{1-a_3}$.

Уравнения сводятся к линейным с помощью логарифмирования. Однако оценки параметров могут оказаться смещенными. В общем случае используется нелинейный МНК, минимизирующий $Q = \sum e_i^2 = \sum \left(y_i - f(a)\right)^2$.

Замечание 9. Модель $y = a_1 x_1^{a_2} x_2^{a_3}$ называют ***мультипликативной*** регрессией, которая также может сводиться к линейной с помощью логарифмирования или же оценивается с помощью частных производных.

21. Уравнения регрессии

21.1. Проверка уравнения регрессии

С помощью простой схемы дисперсионного анализа (полностью случайный план эксперимента) можно проверить наличие связи между переменными. В качестве фактора здесь выступает независимая переменная. Для каждого уровня фактора имеется несколько наблюдений зависимой переменной.

Если систематическая дисперсия значительно выше случайной, то гипотеза об отсутствии связи отклоняется (т. е. принимается наличие связи).

(1) В случае проверки линейной связи сумма квадратов отклонений, $\sum \left(Y_i - \bar{Y}\right)^2 = \underbrace{\sum \left(\widehat{Y}_i - \bar{Y}\right)^2}_{\text{сумма квадратов регрессии}} + \underbrace{\sum e_i^2}_{\text{отклонение от регрессии}}$,

где $e_i = Y_i - \hat{y}_i$.

Схема дисперсионного анализа:

- $Q_1 = \sum e_i^2 = \sum \left(Y_i - \hat{y}_i\right)^2 - \{\nu = n - 2\}$ и $s_1^2 = \dfrac{Q_1}{n-2}$;

- $Q_2 = \sum\left(\hat{y}_i - \overline{Y}\right)^2$ с $\nu = 1$ и $s_2^2 = Q_2$;
- $F = \dfrac{s_2^2}{s_1^2}; \left\{\dfrac{1}{n-2}\right\}$;
- $\underline{\alpha} \xrightarrow{\text{табл.}} F_\alpha;\ \widehat{F} = \dfrac{S_2^2}{S_1^2} > F_\alpha \Rightarrow H_0 : \sigma_1^2 = \sigma_2^2$ — отклоняется, т. е. влияние случайной составляющей несущественно.

(2) В случае проверки гипотезы о нелинейной связи определяются отклонения $\sum e_i^2(\text{лр})$ и $\sum e_i^2(\text{нлр})$.

Их разность $\left(\sum e_i^2(\text{лр}) - \sum e_i^2(\text{нлр})\right)$ показывает, как изменяется случайная составляющая суммы квадратов отклонений. *Схема дисперсионного анализа*:

- $Q_0 = \sum e_i^2(л) = \sum\left(Y_i - \hat{y}_i\right)^2$ c $\{\nu = n-2\}$;
- $Q_1 = \sum e_i^2(\text{нлр}) = \sum\left(Y_i - \hat{y}_i\right)^2$ c $\{\nu = n-3\}$; $s_1^2 = \dfrac{Q_1}{n-3}$ — для параболической регрессии;
- $Q_2 = Q_0 - Q_1$ c $\{\nu = 1\}$ и $s_2^2 = Q_2$;
- $F = \dfrac{s_2^2}{s_1^2}; \left\{\dfrac{1}{n-3}\right\}$;
- $\underline{\alpha} \xrightarrow{\text{по таблице}} F_\alpha; \widehat{F} = \dfrac{s_2^2}{s_1^2} > F_\alpha \Rightarrow H_0 : \sigma_1^2 = \sigma_2^2$ — отклоняется (H_0: изменение случайной составляющей несущественно).

21.2. Структура уравнений регрессии

Определение 1. Под ***структурой*** уравнения регрессии будем понимать количество переменных и взаимных соответствий (связей) между ними.

Как составить уравнение? При отборе переменных для уравнения регрессии стремятся ограничить число учитываемых факторов и оставить только те переменные, которые вносят *ощутимый* вклад в объяснение изменения зависимой переменной.

Приемы составления уравнений регрессии (отбора переменных):

1) Из полного набора переменных в развернутом уравнении регрессии *последовательно исключаются те*, чей вклад в сумму квадратов после оценивания оказывается наименьшим по сравнению с установленным уровнем.
2) Структура уравнения определяется в ходе *«наращивания»* его переменными.

Основная проблема при составлении уравнения регрессии *мультиколлинеарность*, т. е. наличие связи между независимыми переменными в уравнении регрессии.

Если такая связь строгая (например, линейная), то матрица $X^T X$ в уравнении $\vec{a} = \left(X^T X\right)^{-1} X^T \vec{y}$ (п. 20.6) не будет иметь обратной в силу $det X^T X = 0$ и оценки $\vec{a}$ найти не удается. При нестрогой связи $det X^T X \approx 0$ оценки ненадежны, т. е. плохие.

На наличие мультиколлинеарности указывает $det X^T X \cong 0$.

Устранение мультиколлинеарности достигается путем пересмотра структуры уравнения с помощью исключения одной из двух взаимосвязанных переменных или трансформации соответствующих переменных (например, вместо переменных берут их приращения и т. п.).

21.3. Система регрессионных уравнений

При статистическом описании комплекса зависимостей (а не отдельных зависимостей) необходимо разработать систему регрессионных уравнений.

(1) *Взаимозависимая линейная модель.*

Определение 2. ***Эндогенными*** переменными в системе регрессионных уравнений называются переменные, определяемые моделью изучаемого явления, а ***экзогенными*** — независимые от структуры модели (их значение устанавливается вне модели). Модель содержит параметры — коэффициенты, определяемые в ходе статистического оценивания.

Определение 3. Модель системы регрессионных уравнений, в которой зависимые переменные одних уравнений выступают в качестве независимых других, называют ***взаимозависимой***. Взаимозависимая модель имеет две формы: структурную и приведенную.

Приведенная форма есть результат решения структурной.

Структурная форма: $\Gamma\vec{y} + A\vec{x} = \vec{\varepsilon}$, где Γ, A — $(n \times n)$ и $(n \times m)$ — матрицы неизвестных параметров; $\vec{y} - (n \times 1)$ — вектор из эндогенных переменных; $\vec{x} - (m \times 1)$ — вектор из экзогенных переменных; $\vec{\varepsilon}$, $\vec{\varepsilon} \in N(0, \vec{\sigma})$ — случайный вектор.

Приведенная форма:

$\vec{y} = -\Gamma^{-1} A\vec{x} + \Gamma^{-1}\vec{\varepsilon} = B\vec{x} + \vec{\eta}$, где $B \equiv -\Gamma^{-1} A$; $\vec{\eta} \equiv \Gamma^{-1}\vec{\varepsilon}$.

(2) *Оценивание параметров взаимозависимой системы.*
Задача: оценить параметры Γ и A по значениям B из приведенной формулы. Поскольку $B = -\Gamma^{-1}A$, то $B+\Gamma^{-1}A=O_{nxm}$ или $\Gamma B+A=O_{nxm}$. Но задача оценки Γ и A из последнего уравнения неоднозначна. Необходимо вводить ограничения на Γ и A.

1. Если их количество позволяет однозначно определить по B искомые Γ и A, то модель называют ***точно идентифицируемой***. В этом случае используется *косвенный* МНК. Его суть заключается в первоначальной оценке параметров каждого уравнения приведенной формы модели в отдельности с помощью МНК. Затем эти параметры трансформируются в параметры структурной формы.
2. Если количество ограничений на Γ и A больше, то система ***сверхидентифицируемая***. Используется *двухшаговый* МНК. На первом шаге оцениваются параметры приведенной формы с оценкой систематической и случайной составляющей, т. е. предполагается $y_i = \hat{y}_i + v_i$, где $\hat{y}_i$ — оценки по приведенной форме $\hat{y}_i = d_{i1}x_1 + \ldots + d_{ij}x_j$. На втором шаге часть эндогенных переменных заменяется оценками $\hat{y}_i$, далее к структурному уравнению применяется МНК.

22. Введение во временные ряды

22.1. Задачи анализа

Для изучения временных рядов необходимо выбрать подходящую математическую модель. Такой моделью может быть дискретный случайный процесс, строгое определение которого дано в п. 11.1. Дадим эквивалентное определение.

Определение 1. Последовательность случайных величин с согласованным законом распределения называется ***дискретным случайным процессом.***

Определение 2. ***Временным рядом*** называют последовательность реализаций случайных величин. Каждый член (уровень) последовательности связан с соответствующим моментом времени или временным интервалом. Если не наблюдается изменений в средних значениях членов (случайной) последовательности, такой ряд называют ***стационарным.***

Пример (последовательность наблюдений величин во времени).

Основные задачи исследования временных рядов:

1) *Описание изменения* соответствующего показателя (например, среднего) во времени и *выявление свойств ряда* (способы — сглаживающие фильтры; подбор кривых — тренда тенденции; выделение сезонных и других периодических колебаний; расчет среднего темпа роста и т. д.).
2) *Объяснение механизма* изменения уровней временного ряда (с помощью регрессионного анализа).
3) *Статистическое прогнозирование* — экстраполяция тенденции.

22.2. Некоторые приемы выявления тенденции временных рядов

(1) *Фильтрация и сглаживание временных рядов (фильтрация).*

Определение 3. Под ***фильтрацией*** временных рядов понимают операцию замены членов последовательности наблюдений во времени расчетными оценками средних на временном интервале по некоторому правилу (с меньшей изменчивостью).

Замечание 1. Если оценка осуществляется на момент внутри заданного временного интервала, то операцию называют ***сгла-***

живанием, если на момент правого конца интервала — ***фильтрацией***, на последующие моменты за интервалом — ***прогнозом***. Правило, согласно которому ищется оценка, называется ***фильтром***. Наиболее употребительны *линейные фильтры*, имеющие вид $\hat{y}_t = \sum_{r=-q}^{s} a_r y_{t+r}$, где a_r — весовой коэффициент; y_{t+r} — наблюдение на момент $(t+r)$; q — уровни (№ значений) до момента t; s — уровни (номера значений) после момента t.

(2) *Скользящие средние*.
Если $\sum a_r = 1$ и $a_r = const$, то фильтр будет соответствовать средней арифметической.

Определение 4. Среднеарифметическую оценку на скользящем временном интервале с единичным шагом называют ***скользящей средней***.

Если принять $a_r = \dfrac{1}{2p+1}$, где $r = -p, . \geq , p$; то $\hat{y}_{t_{cc}} = \dfrac{1}{2p+1} \sum_{i=t-p}^{t+p} y_i$.

Замечание 2. Чаще всего на практике $m = 2p+1$ принимают 3, 5, 7 лет.

Замечание 3. При $m > 3$ можно пользоваться рекуррентной формулой:

$$\hat{y}_t = \hat{y}_{t-1} + \frac{y_{t+p} - y_{t-(p+1)}}{2p+1}.$$

Очевидно $D\left(\hat{y}_t\right) = \sigma^2 / m$.

Скользящая средняя является довольно грубой обработкой.

Замечание 4. Более тонкой модификацией является *взвешенная скользящая средняя*, когда в пределах интервала каждому значению y_i приписывается вес в зависимости от расстояния от i до середины интервала.

Если скользящая средняя заменяет не центральный, а последний член в интервале сглаживания, то $\hat{y}_t = \dfrac{1}{m} \sum_{r=0}^{m} y_{t-r}$.

Откуда можно получить $\hat{y}_t = \hat{y}_{t-1} + \dfrac{y_t - y_{t-m}}{m}$.

Степень «обновления» средней определяется весом $\frac{1}{m}$ в последнем члене.

(3) *Экспоненциальная средняя*: $\hat{y}_t = \sum_{r=0}^{t} \alpha(1-\alpha)^r y_{t-r}$, где $m = t$, α — коэффициент, характеризующий вес текущего наблюдения (параметр сглаживания: $0 < \alpha < 1$). Из последней формулы с помощью перегруппировки легко получить $\hat{y}_t = \hat{y}_{t-1} + \alpha\left(y_t - \hat{y}_{t-1}\right)$. Обычно $\alpha \approx 0{,}2$ или α берут из интервала [0,1; 0,3]. Эта формула обычно используется для прогноза.

Нетрудно показать, что $D\left(\hat{y}_{t_{\text{экс}}}\right) = \frac{\alpha}{2-\alpha}\sigma^2$.

(4) *Метод последовательных разностей.*

$$y_t = \overline{y}_t + \varepsilon_t,$$

где $\overline{y}_t$ — систематическая (регулярная) компонента; ε_t — случайная компонента, $\overline{y}_t = f(t)$ определяется, например, полиномом. Полагают $\varepsilon_t \in N(0, \sigma^2)$ и $E\varepsilon_t\varepsilon_r = 0$. Далее вычисляются первые разности $U_t = y_t - y_{t-1}$; вторые разности $U_t^{(2)} = U_t - U_{t-1}$ и так далее, пока разности не будут примерно равными. Порядок разности принимается за степень полинома.

22.3. Средний темп роста

Он характеризует изменение ряда и определяется как геометрическая средняя из последовательных (цепных) темпов роста. Пусть имеем $y_1, y_2, \ldots, y_n$.

Темп роста есть $\tau_i = \frac{y_{i+1}}{y_i}$, $i = 1, 2, \ldots, n-1$, а средний темп

$$\overline{\tau} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[n-1]{\tau_1 \cdot \tau_2 \ldots \tau_{T-1}}.$$

Тогда для прогноза на момент $m > n$, имеем: $y_m = y_1\left(\overline{\tau}\right)^{m-1}$.

23. Некоторые специальные методы многомерного статистического анализа

23.1. Основные понятия и задачи многомерного статистического анализа

Многомерный статистический анализ (МСА) является специальным разделом прикладной статистики. Его основная цель состоит в выявлении характера и структуры взаимосвязей между компонентами многомерного признака. МСА можно рассматривать и как составную часть общего регрессионного анализа. Данные о социально-экономических объектах и процессах обычно представляются в виде многомерных наблюдений, например, в виде матрицы:

$$\begin{pmatrix} X_{11}(t) & X_{12}(t) & \dots & X_{1m}(t) \\ X_{21}(t) & X_{22}(t) & \dots & X_{2m}(t) \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1}(t) & X_{n2}(t) & \dots & X_{nm}(t) \end{pmatrix},$$

где $X_{ij}(t)$ — значение j-го анализируемого признака, характеризующего состояние i-го объекта в момент времени t. Если наблюдения относятся к одному моменту времени, то индекс t опускается. Многомерный признак в общем случае можно записать в виде вектор-строки $(x_1(t),\ x_2(t),\ \dots,\ x_m(t))$, или, опуская индекс t, он записывается как вектор-столбец $\vec{x} = (x_1,\ x_2,\ \dots,\ x_m)^{\text{т}}$.

Реальный объект (процесс, явление, система), как правило, представляется набором входных переменных, т. н. *экзогенных* (объясняющих), и выходных переменных, т. н. *эндогенных* (объясняемых). Многомерный признак обычно и описывается этими двумя наборами. Среди его компонент могут быть количественные, порядковые (ординальные) и классификационные (номинальные). Если природа результирующих (эндогенных) показателей количественная, то для исследования обычно используют такие разделы МСА как регрессионный и корреляционный анализ, анализ временных рядов, дисперсионный анализ. Если же природа этих показателей носит неколичественный характер, например, порядковый или классификационный, то применяют анализ ранговых корреляций, дискриминантный и кластер-анализ. Очевидно, что наряду с традиционными задачами оценивания и проверки гипотез в МСА приобретают важную роль задачи снижения размерности и классификации многомерного признака.

Число составляющих (компонент) признака m может быть очень большим. Это само по себе создает большие трудности при реализации процедур обработки исходных данных. В то же время некоторые составляющие-признаки могут быть взаимосвязаны, т. е. информация дублируется; ряд признаков оказываются неинформативными, при переходе от одного объекта к другому они практически не меняются; некоторые признаки можно «суммировать». Все это дает основание поставить задачу снижения размерности многомерного признака. При ее решении наибольшее распространение получили методы *факторного анализа,* метод *главных компонент*.

23.2. О моделях и методах факторного анализа в МСА

Структура связей между m составляющими признака $(x_1, x_2, \geq, x_m)$ может объясняться тем, что эти переменные зависят от меньшего числа других, неизмеряемых («скрытых») факторов $(f_1, f_2, \geq, f_r)$ $r < m$, некоррелированных между собой. Основная цель факторного анализа — выявление скрытых общих факторов с минимизацией их числа и степени зависимости составляющих признака от остальных случайных компонент. Общие факторы можно считать причинами, а наблюдаемые признаки следствиями. Основным предположением является взаимная независимость (некоррелированность) исходных составляющих признака. Пусть имеются центрированные наблюдения $X_1, X_2, \geq, X_n$ (любое наблюдение центрируется вычитанием из него среднего). Тогда *линейная* (наиболее распространенная) модель факторного анализа имеет вид $\vec{x} = Q\vec{f} + \vec{\varepsilon}$, где $\vec{x}$ — есть m-мерный признак; $Q = \left\{q_{ij}\right\}_{\substack{i=1,2,\ldots,m \\ j=1,2,\ldots,r}}$ — матрица «нагрузок» общих факторов на исследуемые признаки; $\vec{f} = \left(f_1, f_2, \ldots, f_r\right)^{\text{т}}$ — вектор факторов; $\vec{\varepsilon} = \left(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m\right)^{\text{т}}$ — вектор остаточных случайных компонент.

Для конкретного наблюдения $X_\nu(\nu = 1,\ 2,\ \ldots,\ n)$ имеем

$\vec{X}_\nu = Q\vec{f}_\nu + \vec{\varepsilon}_\nu$.

При допущении о нормальном характере и некоррелированности остаточных случайных компонент, т. е.

$\vec{\varepsilon} \in N(0, V);\ V = E\vec{\varepsilon} \cdot \vec{\varepsilon}^{\,\text{т}} = \{diagD\varepsilon_i\}_{i=1,2,\ldots,m}$, и поскольку

$E\vec{x} = 0,\ E\left(\vec{x} \cdot \vec{x}^{\,\text{т}}\right) = R = QQ^{\text{т}} + V,$ то можно считать

$E\vec{f} = 0,\ E\vec{f}\vec{f}^{\,\text{т}} = I_r$ и $\vec{f} \in N(0, I_r)$.

Замечание 1. Линейная модель факторного анализа имеет формальное сходство с моделью множественной регрессии. Отличие состоит в том, что в модели факторного анализа переменные $\vec{f}$ не являются непосредственно наблюдаемыми, тогда как в модели множественной регрессии они являются аргументами.

Основными задачами факторного анализа являются:

- задача существования модели (каково соотношение между r и m и при каких условиях справедливо линейное представление);
- задача единственности модели (при каких ограничениях на матрицу Q и ковариационную матрицу V определение параметров модели $\vec{f}$ и Q единственно);
- как конкретно вычислить неизвестные параметры модели;
- статистическое оценивание неизвестных структурных параметров модели (по наблюдениям при заданном числе факторов r);
- статистическая проверка гипотез (об истинном числе r и неравенстве нулю элементов матрицы Q);
- построение статистических оценок для ненаблюдаемых факторов $\vec{f}_i,\ i = 1,\ 2,\ \ldots,\ n$.

При решении указанных задач используются различные приемы и методы (максимального правдоподобия, центроидный метод для статистического оценивания факторных «нагрузок» и остаточных дисперсий; метод Бартлетта и метод Томсона при оценке значений общих факторов).

Кроме того, существуют также частные случаи логической схемы факторного анализа на основе эвристического подхода: метод экстремальной группировки параметров, метод корреляционных плеяд.

23.3. Метод главных компонент

Пусть размерность признака при обследовании объектов есть m и необходимо перейти к новому признаку с меньшей размерностью r. Это значит, что с помощью вектора $\vec{x}$ надо найти новый вектор $\vec{z} = (z_1, z_2, \ldots, z_r)^{\text{т}}$ вспомогательных показателей с возможно малой размерностью ($r = 1, 2, 3$).

Новые признаки $z_1, z_2, \ldots, z_r$ могут выбираться из исходных по какому-либо правилу, например, как линейные комбинации исходных.

Пусть имеется анализируемый m-мерный признак $\vec{x}$ случайной величины с вектором средних $\vec{a} = (a_1, a_2, \ldots, a_m)^{\text{т}}$ и ковариационной (корреляционной) матрицей $R = \{\sigma_{ij}\}_{\substack{i=1,2,\ldots,m \\ j=1,2,\ldots,m}}$.

Введем класс допустимых преобразований исследуемых признаков $x_1, x_2, \geq, x_m$ в виде линейных ортогональных нормированно-центрированных комбинаций

$$F = \left\{ \vec{z} : z_j = \sum_{\nu=1}^{m} c_{j\nu}(x_\nu - a_\nu);\ j = 1,\ 2,\ \ldots,\ r \right\},$$

где $\sum c_{j\nu}^2 = 1$, $\sum c_{j\nu} c_{k\nu} = 0$ для $j = 1, 2, \ldots, r$ и $k = 1, 2, \ldots, r$; $j \neq k$.

В качестве меры информативности обычно принимают

$$I_r(\vec{z}(\vec{x})) = \frac{Dz_1 + Dz_2 + \cdots + Dz_r}{Dx_1 + Dx_2 + \cdots + Dx_m},$$

где D — есть оператор (символ) дисперсии.

Тогда при любом фиксированном r вектор новых признаков $\vec{z}(\vec{x}) = (z_1(\vec{x}), z_2(\vec{x}), \ldots, z_r(\vec{x}))^{\text{T}}$ определяется как линейная комбинация $\vec{z} = L\vec{x}$, где $L = \begin{pmatrix} l_{11} & \cdots & l_{1m} \\ \vdots & \vdots & \vdots \\ l_{r1} & \cdots & l_{rm} \end{pmatrix}$ есть матрица, строки которой удовлетворяют условию ортогональности, и когда

$I_r(z_1(\vec{x}),\ z_2(\vec{x}),\ \ldots,\ z_r(\vec{x})) = \max I_r(\vec{z}(\vec{x}))$ по всем $\vec{z}(\vec{x}) \in F$.

Новые переменные $z_1(\vec{x})$, $z_2(\vec{x})$, …, $z_r(\vec{x})$ называются *главными компонентами* вектора $\vec{x}$.

Определение 1. *Первой главной компонентой* $z_1(\vec{x})$ исследуемой системы показателей $\vec{x} = (x_1, x_2, \ldots, x_m)^{\mathrm{T}}$ называется такая линейная (нормированная и центрированная) комбинация этих показателей, которая среди всех прочих линейных комбинаций такого типа обладает наибольшей дисперсией.

Определение 2. *k-й главной компонентой* $z_k(\vec{x})$ системы показателей называется линейная (некоррелированная с предыдущими главными компонентами, нормированная и центрированная) комбинация с наибольшей дисперсией среди линейных комбинаций такого типа.

Вид матрицы L зависит только от элементов корреляционной (ковариационной) матрицы R. Исходную систему показателей обычно предварительно центрируют, и тогда можно считать $Ex_j = 0$; $j = 1, 2, \ldots, m$.

Оптимизационная задача на условный экстремум имеет вид

$$\begin{cases} D\left(\vec{l}_1^{\mathrm{T}} \vec{x}\right) \to \max \\ \vec{l}_1^{\mathrm{T}} \cdot \vec{l}_1 = 1 \end{cases},$$

где $\vec{l}_1$ — вектор из элементов первой строки в матрице L. Решая задачу по методу Лагранжа, получаем систему уравнений для ее определения:

$$D\left(\vec{l}_1^{T} \vec{x}\right) = E\left(\vec{l}_1^{T} \vec{x}\right)^2 = E\left(\vec{l}_1^{T} \vec{x} \cdot \vec{x}^{T} \vec{l}_1\right) = \vec{l}_1^{T} R \vec{l}_1; \ \varphi\left(\vec{l}_1, \lambda\right) = \vec{l}_1^{T} R \vec{l}_1 - \lambda\left(\vec{l}_1^{T} \vec{l}_1 - 1\right);$$

$$\frac{\partial \varphi}{\partial \vec{l}_1^{T}} = 2R\vec{l}_1 - 2\lambda \vec{l}_1 = 0, \ \left(R - \lambda I\right)\vec{l}_1 = 0.$$

Для существования ненулевого решения необходимо

$$det\left(R - \lambda I\right) = 0.$$

При симметричности и неотрицательной определенности матрицы R это уравнение имеет m вещественных неотрицательных корней — собственных чисел $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_m \geq 0$. Поскольку $Dz_1 = D\left(\vec{l}_1^{\mathrm{T}} \vec{x}\right) = \vec{l}_1^{\mathrm{T}} R \vec{l}_1$, а из системы имеем $\vec{l}_1^{\mathrm{T}} R \vec{l}_1 = \lambda$, то $Dz_1(\vec{x}) = \lambda$.

Для обеспечения максимальной дисперсии переменной z_1 нужно выбрать наибольшее из собственных чисел матрицы R, т. е.

$Dz_1(\vec{x}) = \lambda_1$. Это собственное число внесем в уравнение $(R - \lambda_1 I)\vec{l}_1 = 0$, откуда находим собственный вектор $\vec{l}_1$.

Таким образом, 1-я главная компонента есть линейная комбинация $z_1(\vec{x}) = \vec{l}_1^{\,\mathrm{T}}\vec{x}$. Аналогично получают и другие главные компоненты $z_k(\vec{x}) = \vec{l}_k^{\,\mathrm{T}}\vec{x}$. Матрица L состоит из строк

$$\vec{l}_j^{\,\mathrm{T}} = (l_{j1},\ l_{j2},\ \ldots,\ l_{jm});\ j = 1,\ 2,\ \ldots,\ r,$$

которые являются собственными векторами матрицы R.

Для главных компонент имеют место следующие свойства:

1) математическое ожидание главных компонент равно нулю

$$E\vec{z} = E(L\vec{x}) = L(E\vec{x}) = 0;$$

2) корреляционная матрица главных компонент есть

$$R_z = E(\vec{z}\vec{z}^{\,\mathrm{T}}) = L\left(E(\vec{x}\vec{x}^{\,\mathrm{T}})\right)L^{\mathrm{T}} = LRL^{\mathrm{T}},$$

$$\vec{l}_j^{\,\mathrm{T}}(R - \lambda_k I)\vec{l}_k = \vec{l}_j^{\,\mathrm{T}} \cdot 0,\ \vec{l}_j^{\,\mathrm{T}} R\vec{l}_k = \lambda_k \vec{l}_k,\ (k = 1,\ 2,\ \ldots,\ m;\ j = 1,\ 2,\ \ldots,\ r).$$

Откуда следует $R_z = LRL^{\mathrm{T}} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_r \end{pmatrix}$;

3) сумма дисперсий исходных признаков равна сумме дисперсий главных компонент $\sum_{k=1}^{m} Dx_k = \sum_{k=1}^{m} Dz_k$.

Из указанных свойств следует, что в качестве меры информативности для метода главных компонент можно взять

$$I_r(\vec{z}(\vec{x})) = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_r}{\lambda_1 + \lambda_2 + \cdots + \lambda_m}.$$

Для нормированных признаков $\left(D\left(x_j\right)=1\right)$ имеем

$$I_r\left(\vec{z}\left(\vec{x}\right)\right)=\frac{\lambda_1+\lambda_2+\cdots+\lambda_r}{m}.$$

Для выборочных характеристик $\hat{\lambda}_i$ и $\hat{\vec{l}}_i$ (по выборочной матрице $\hat{R}$) при условии $\vec{x}_i \in N\left(0,R\right)$, $i=1,\ 2,\ \ldots,\ m$ известно, что как оценки максимального правдоподобия они являются оценками состоятельными, несмещенными и асимптотически эффективными. Для них можно находить и интервальные оценки.

23.4. Классификация объектов. Элементы кластер-анализа

Определение 3. Под *классификацией* понимают 1) разделение рассматриваемой совокупности объектов или явлений на однородные (в некотором смысле) группы либо 2) включение каждого объекта из множества к какому-то заранее известному классу.

Замечание 1. Если кроме классифицируемой совокупности имеются выборочные данные по каждому классу, то задачу классификации называют *классификацией с обучением*, если же таких «обучающих» выборок не имеется, то задачу называют *классификацией без обучения*.

Замечание 2. Задача классификации, например, предшествует применению множественной регрессии. Все имеющиеся данные должны быть разбиты на однородные классы, после чего ищутся коэффициенты регрессии для каждого класса.

Замечание 3. Исходные данные при решении задачи классификации могут быть представлены не только в виде матрицы наблюдений X (см. п. 23.1), но и в виде матрицы парных сравнений

$$\Gamma=\begin{pmatrix} \gamma_{11}(t) & \gamma_{12}(t) & \cdots & \gamma_{1m}(t) \\ \gamma_{21}(t) & \gamma_{22}(t) & \cdots & \gamma_{2m}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1}(t) & \gamma_{m2}(t) & \cdots & \gamma_{mm}(t) \end{pmatrix}.$$

Обычно рассматриваются одномоментные сравнения, поэтому индекс t можно опустить. В матрице парных сравнений принимают $m = n$.

Общая постановка задачи классификации состоит в следующем: всю анализируемую совокупность объектов, статистически представимую через матрицу наблюдений (матрицу парных сравнений), необходимо разделить на сравнительно небольшое число (заранее известное или нет) однородных в определенном смысле групп или классов.

В зависимости от исходной информации о природе выявляемых классов и поставленной задачи используют либо 1) методы «расщепления» смесей вероятностных распределений (каждый класс рассматривается как одномодальная генеральная совокупность с плотностью $p_j\left(x,\theta_j\right)$, $j=1,\ 2,\ \ldots,\ k$ и неизвестным параметром θ_j, а каждое наблюдение X_i считается извлеченным из одной из этих генеральных совокупностей, неизвестно из какой именно, либо 2) методы автоматической классификации (кластер-анализа), когда нет достаточной информации для параметрического представления искомых классов.

1) Методы «расщепления» составляют основу *дискриминантного* анализа, в котором предполагается наличие обучающих выборок. При дискриминантном анализе наблюдение относят к тому классу (той генеральной совокупности), в рамках которого (которой) оно выглядит более правдоподобно. Исходными для классификации являются m-мерные наблюдения X_1, X_2, $\geq X_n$ и k обучающих выборок. Общая генеральная совокупность в виде смеси k классов имеет плотность распределения $p(x)=\sum_{j=1}^{k} P_j p_j\left(x,\theta_j\right)$, где P_j — вероятность появления в выборке элемента из класса j с плотностью $p_i(x)$ или «удельный вес» элементов j-го класса во всей генеральной совокупности; θ_j — неизвестные параметры плотностей соответствующих классов.

Решить задачу «расщепления» смеси распределений — это значит по выборке классифицируемых наблюдений X_1, X_2, $\geq X_n$ из общей генеральной совокупности построить статистические оценки для числа компонентов смеси k, их удельных весов P_i и для каждого из компонентов $p_j\left(x,\theta_j\right)$ анализируемой смеси. Можно рассматривать частные случаи, когда известны k и(или) P_j, а оценивается $p_j\left(x,\theta_j\right)$ с предварительным оцениванием по обучающим выборкам векторного параметра θ_j; или же известны $p_j(x)$ и по ним оцениваются P_j.

Пусть имеется наблюдение X_i. К какой из k анализируемых генеральных совокупностей $p_j(x, \theta_j)$, $j = 1, 2, \geq, k$ следует его отнести? Для решения этой задачи сначала оценивают неизвестный параметр θ_j, используя обучающие выборки, затем поочередно вычисляют значения функций правдоподобия для имеющегося наблюдения X_i в рамках каждой генеральной совокупности $p_j\left(x_{(1)}, x_{(2)}, \ldots, x_{(n)}\right)$ и относят это наблюдение к тому классу, для которого функция правдоподобия оказывается наибольшей.

2) Методы автоматической классификации — *кластер-анализа* — применяются при отсутствии информации о вероятностном описании классов. В этом случае вводится «признаковое» пространство, в котором объекты рассматриваются в виде точек. Такими точками в m-мерном пространстве являются

$$\vec{X}_1 = \left(X_{11}, X_{12}, \ldots, X_{1m}\right)^{\text{т}},$$

$$\vec{X}_2 = \left(X_{21}, X_{22}, \ldots, X_{2m}\right)^{\text{т}}, \ldots, \vec{X}_n = \left(X_{n1}, X_{n2}, \ldots, X_{nm}\right)^{\text{т}}.$$

Обучающие выборки отсутствуют. Близость между объектами можно принимать за однородность. Тогда классификация будет заключаться в расчленении совокупности точек-наблюдений на классы так, чтобы объекты в классе находились на небольших расстояниях друг от друга. Эти классы называют *кластерами* или *образами*.

За количественную характеристику однородности (меру близости) принимают специальную метрику в пространстве признаков-объектов. Это расстояние сопоставляется с некоторым порогом, определяемым в каждом случае целями задачи. Если наблюдения $\vec{X}_i$ извлекаются из нормальных генеральных совокупностей с одной и той же ковариационной (корреляционной) матрицей R, то за меру близости двух объектов удобнее всего принять *расстояние Механобиса* $d_M\left(\vec{X}_i, \vec{X}_j\right) = \sqrt{\left(\vec{X}_i - \vec{X}_j\right)^{\text{т}} \Lambda^{\text{т}} R^{-1} \Lambda \left(\vec{X}_i - \vec{X}_j\right)}$, где R — ковариационная (корреляционная) матрица общей генеральной совокупности, из которой извлекаются наблюдения X_i; Λ — симметричная, неотрицательно определенная матрица весовых коэффициентов λ_{mr} (обычно диагональная). Частными случаями этой метрики являются:

– *евклидово* расстояние $d_E\left(\vec{X}_i,\vec{X}_j\right)=\sqrt{\sum_{k=1}^{m}\left(X_{ik}-X_{jk}\right)^2}$.

  Оно используется при многомерной нормальной генеральной совокупности с ковариационной матрицей R (т. е. компоненты X_{ij} взаимно независимы и имеют одну и ту же дисперсию), при условии однородности по физическому смыслу компонент вектора наблюдений $\vec{X}_i=\left(X_{i1},X_{i2},\ldots,X_{im}\right)^{\text{т}}$ и при размерности пространства признаков $m = 1, 2, 3$.

– *взвешенное евклидово* расстояние

  $d_{вE}\left(\vec{X}_i,\vec{X}_j\right)=\sqrt{\sum_{k=1}^{m}w_k\left(X_{ik}-X_{jk}\right)^2}$, где w_k — «вес» (коэффициент важности), $w_k\geq 0$, $\sum_{k=1}^{m}w_k=1$.

– расстояние *Хемминга* $d_H\left(\vec{X}_i,\vec{X}_j\right)=\sum_{k=1}^{m}\left|X_{ik}-X_{jk}\right|$.

В ряде ситуаций при классификации оказывается удобным вводить понятие расстояния между группами объектов. Пусть S_l есть l-я группа (кластер) объектов, n_l — число объектов, образующих группу S_l, вектор $\vec{X}(l)$ — среднее арифметическое векторных наблюдений, входящих в S_l («центр тяжести» l-й группы). Тогда расстояние между группами обозначают через $\rho\left(S_l,S_q\right)$. Наиболее употребительными расстояниями между группами S_l и S_q являются:

1) «ближнего соседа»

   $\rho_{\min}\left(S_l,S_q\right)=\min d\left(\vec{X}_i,\vec{X}_j\right)$ по всем $\vec{X}_i\in S_l,\vec{X}_j\in S_q$;

2) «дальнего соседа»

   $\rho_{\max}\left(S_l,S_q\right)=\max d\left(\vec{X}_i,\vec{X}_j\right)$ по всем $\vec{X}_i\in S_l,\vec{X}_j\in S_q$;

3) по «центрам тяжести» $\rho\left(S_l, S_q\right) = d\left(\vec{X}(l), \vec{X}(q)\right)$;

4) по «средней связи» $\rho_{cp}\left(S_l, S_q\right) = \dfrac{1}{n_l \cdot n_q} \sum\limits_{\vec{X}_i \in S_l} \sum\limits_{\vec{X}_j \in S_q} d\left(\vec{X}_i, \vec{X}_j\right)$.

Для определения качества расчленения заданной совокупности на классы используются различные функционалы. Например, сумма внутриклассных дисперсий (k классов):

$$Q(S) = \sum_{l=1}^{k} \sum_{\vec{X}_i \in S_l} d^2\left(\vec{X}_i, \vec{X}(l)\right).$$

23.5. Применение методов МСА для анализа временных рядов (метод «гусеница»)

Этот метод применяется для обработки временных рядов. Основная идея метода заключается в том, что временной ряд с помощью операции расчленения представляют в виде многомерных исходных данных, к которым применяют метод главных компонент, а затем восстанавливают исходный ряд.

Пусть имеется временной ряд $\left\{x_i\right\}_{i=1}^{N}$, образованный последовательностью N равноотстоящих значений некоторой (возможно, случайной) функции.

Выбирается некоторое число $M<N$, называемое *длиной гусеницы,* и представляются первые M значений последовательности в качестве первой строки матрицы наблюдений. В качестве второй строки матрицы берутся значения последовательности с x_2 по x_{M+1}. Последней строкой с номером $k = N - M + 1$ будут последние M элементов последовательности. Тогда имеем

$$X = \left\{x_{ij}\right\}_{i,j=1}^{k,\,M} = \begin{pmatrix} x_1 & x_2 & \cdots & x_M \\ x_2 & x_3 & \cdots & x_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k+1} & \cdots & x_N \end{pmatrix}.$$

Эту матрицу, элементы которой равны $x_{ij} = x_{i+j-1}$, можно рассматривать как M-мерную выборку объема k. Матрица записана в виде «строка — индивид, столбец — признак».

Далее по обычной схеме проводится анализ главных компонент.

Процедура восстановления одномерного ряда основана на умножении матрицы главных компонент на ортогональную матрицу L^T. В результате восстанавливаемая матрица представляется в виде суммы M матриц, каждая из которых порождена одним собственным вектором ковариационной матрицы R.

Рекомендуемая литература

1. *Айвазян С. А., Мхитарян В. С.* Прикладная статистика и эконометрика. — М.: ЮНИТИ, 1998.
2. *Бочаров П. П., Печенкин А. В.* Теория вероятностей и математическая статистика. — М.: Гардарика, 1998.
3. *Вентцель Е. С.* Теория вероятностей. — М.: Наука, 1976.
4. *Вентцель Е. С.* Исследование операций: задачи, примеры, методология. — М.: Наука, 1980.
5. *Гмурман В. Е.* Руководство к решению задач по теории вероятностей и математической статистике. — М.: Высшая школа, 1998.
6. *Грешилов А. А., Стакун В. А., Стакун А. А.* Математические методы построения прогнозов. — М.: Радио и связь, 1997.
7. *Дубров А. М., Мхитарян В. С., Трошин Л. И.* Многомерные статистические методы. — М.: Финансы и статистика, 1998.
8. *Ковалев В. А., Калинина В. Н.* Теория вероятностей и математическая статистика. — М.: ИНФРА-М, 1999.
9. *Колде Я. К.* Практикум по теории вероятностей и математической статистике. — М.: Высшая школа, 1991.
10. *Крамер Г.* Математические методы статистики. — М.: Мир, 1975.
11. *Красс М. С.* Математика для экономических специальностей. — М.: ИНФРА-М, 1998.
12. *Кремер Н. Ш.* Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2000.
13. *Гнеденко Б. В.* Курс теории вероятностей. — М.: Наука, 1979.
14. *Гнеденко Б. В., Хинчин А. Я.* Элементарное введение в теорию вероятностей. — М.: Наука, 1970.
15. *Смирнов Н. В., Дунин-Барковский И. В.* Курс теории вероятностей и математической статистики для технических приложений. — М.: Наука, 1965.
16. *Четыркин Е. М., Калихман И. Л.* Вероятность и статистика. — М.: Финансы и статистика, 1982.

Приложение 1

Расчетно-графическая работа

Задание к расчетно-графической работе

Выборки сделаны из генеральной совокупности, распределенной по нормальному закону:

Выборка 1 — заданная статистическая совокупность.

Выборка 2 — первые 25 элементов совокупности.

Выборка 3 — последние 20 элементов совокупности.

1. Для заданной статистической совокупности:
   - составить интервальный вариационный ряд;
   - вычислить относительные частоты;
   - вычислить эмпирическую функцию распределения;
   - построить графики (гистограммы) относительных частот и эмпирической функции распределения;
   - вычислить выборочные: среднее значение, дисперсию, среднеквадратическое отклонение и определить выборочные моду и медиану.
2. Используя выборки 2 и 3, по дискретному вариационному ряду вычислить несмещенные оценки для среднего значения, дисперсии, среднеквадратического отклонения генеральной совокупности.
3. Для выборки 1, считая, что дисперсия генеральной совокупности известна $\sigma^2=T^2$:
   - определить доверительный интервал для оценки среднего значения при доверительной вероятность $P=1-\alpha_1$;
   - по предельной ошибке выборки ε для среднего значения найти соответствующую ему доверительную вероятность;
   - определить необходимый объем выборки для определения среднего значения генеральной совокупности с доверительной вероятностью $P=1-\alpha_2$ и предельной ошибкой выборки ε.
4. Используя выборку 2, определить доверительный интервал, соответствующий доверительной вероятности $P=1-\alpha_3$ для оценки среднего значения генеральной совокупности.
5. Используя выборку 3, определить доверительный интервал, соответствующий доверительной вероятности $P=1-\alpha_4$, для оценки дисперсии генеральной совокупности.
6. Определить доверительный интервал, соответствующий доверительной вероятности $P=1-\alpha_5$, для оценки доли признака.

Объем выборки равен 100, выборочная доля признака по данным наблюдения равна $15+5K$.

7. Проверить по выборке 2 гипотезу о том, что среднее значение генеральной совокупности равно A на уровне значимости α_6 при альтернативной гипотезе — среднее значение не равно A.
8. Проверить по выборке 3 гипотезу о том, что дисперсия генеральной совокупности равна T^2 на уровне значимости α_7 при альтернативной гипотезе — дисперсия не равна T^2.
9. По выборкам 2 и 3 проверить гипотезу о том, что средние значения соответствующих генеральных совокупностей равны на уровне значимости α_8 при альтернативной гипотезе — они не равны.
10. По выборке 1 проверить гипотезу о том, что генеральная совокупность имеет нормальное распределение с параметрами $\bar{X} = A$, $\sigma^2 = T^2$ на уровне значимости α_9.
11. С помощью метода крайних точек («натянутой нити») найти линейные функции регрессии для связанных выборок XY, XZ, YZ и построить их графики.
12. По методу наименьших квадратов найти линейные функции регрессии для двумерных выборок задачи 11 и наложить их графики на графики предыдущей задачи.
13. Полагая, что каждый i-й столбец в выборке 2 соответствует i-му уровню фактора A_i, оценить влияние фактора A на уровне значимости α_{10} при полностью случайном плане эксперимента.
14. Произвести сглаживание заданного временного ряда методом скользящих средних, выявить тренд с использованием метода наименьших квадратов и найти прогноз на 3 временные единицы. Построить графики.

Образец

СЕВЕРО-ЗАПАДНАЯ АКАДЕМИЯ
ГОСУДАРСТВЕННОЙ СЛУЖБЫ

Кафедра математики

РАСЧЕТНО-ГРАФИЧЕСКАЯ РАБОТА
по математической статистике

Выполнил: ***студент 121 группы***
Ковалев Е. А.
Проверил: ***Петров И. А.***

Санкт-Петербург
200_

Исходные данные

1. Генеральная совокупность с нормальным распределением; среднее значение $A = 16$; дисперсия $T^2 = 16$.
2. Предельная ошибка выборки $\varepsilon = 0,65$.
3. Уровни значимости: $\alpha_1 = 0,03$; $\alpha_2 = 0,02$; $\alpha_3 = 0,05$; $\alpha_4 = 0,03$; $\alpha_5 = 0,05$; $\alpha_6 = 0,03$; $\alpha_7 = 0,05$; $\alpha_8 = 0,02$; $\alpha_9 = 0,05$; $\alpha_{10} = 0,05$. $K=0$.

Выборка 1

9,27	12,18	18,91	16,02	14,31
15,51	11,61	18,85	19,79	13,77
15,84	10,24	17,17	20,19	14,90
18,64	15,40	19,71	18,68	15,88
12,73	20,68	19,96	14,60	16,09
16,32	16,20	14,24	15,26	16,12
14,72	11,13	16,50	19,15	10,85
11,75	13,79	19,07	16,27	19,17
16,25	13,90	14,21	17,46	16,88
14,69	16,82	10,35	17,01	16,78
15,22	19,17	19,22	15,36	17,29
20,11	8,54	18,38	15,07	16,73
18,43	16,20	13,13	14,85	11,22
12,48	14,42	19,90	12,37	17,52
14,44	14,32	17,86	15,22	13,27
16,89	20,62	16,42	12,54	13,02
19,32	11,71	16,33	14,48	20,85
9,63	11,23	14,62	15,70	19,98
15,93	15,17	15,88	18,80	17,77
12,12	13,82	19,02	15,62	21,11
14,23	13,66	12,15	21,05	15,02
10,82	10,86	21,71	16,69	17,91
17,50	20,82	18,70	16,85	13,82
16,44	15,34	18,71	12,78	16,44

Выборка 2

9,27	16,32	15,22	16,89	14,23
15,51	14,72	20,11	19,32	10,82
15,84	11,75	18,43	9,63	17,50
18,64	16,25	12,48	15,93	16,44
12,73	14,69	14,44	12,12	12,18

Выборка 3

16,09	16,88	11,22	20,85	15,02
16,12	16,78	17,52	19,98	17,91
10,85	17,29	13,27	17,77	13,82
19,17	16,73	13,02	21,11	16,44

Связанные выборки

X	Y	Z
9,27	5,17	21,11
9,63	6,07	21,05
10,24	6,56	20,85
10,82	7,10	19,98
11,61	7,12	19,17
11,75	7,30	18,80
12,12	7,31	17,91
12,18	7,63	17,77
12,48	7,94	17,52
12,73	8,01	17,29
14,23	8,17	16,88
14,44	8,21	16,85
14,69	8,25	16,78
14,72	8,54	16,73
15,22	8,93	16,69
15,40	9,19	16,44
15,51	9,34	16,12
15,84	9,35	16,09
15,93	9,36	15,88
16,20	9,43	15,62
16,25	9,48	15,02
16,32	9,51	14,90
16,44	9,53	14,31
16,89	9,61	13,82
17,50	9,86	13,77
18,43	9,89	13,27
18,64	9,95	13,02
19,32	9,98	12,78
20,11	10,09	11,22
20,68	10,86	10,85

Временной ряд

Время	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Параметр	42,9	32,8	35,1	40,9	38,9	38,2	42,9	41,5	39,3	47,3	45,8	47,4
Время	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Параметр	44,5	54,0	49,0	44,8	52,4	55,4	56,3	55,2	48,7	57,6	60,9	58,4

1. Для заданной статистической совокупности:
- составить интервальный вариационный ряд;
- вычислить относительные частоты;
- вычислить эмпирическую функцию распределения;
- построить графики (гистограммы) относительных частот и эмпирической функции распределения;
- вычислить выборочные: среднее значение, дисперсию, среднеквадратическое отклонение и определить выборочные моду и медиану.

Объем выборки n=120; $X_{\min}$=8,54; $X_{\max}$=21,7; размах R=$X_{\max}$–$X_{\min}$=21,71–8,54=13,17; количество интервалов $N \approx \sqrt{n} \sim 10$; длина интервала k=1,5; C=16,0 (соответствует середине интервала x_i для максимальной частоты m_i).

Интервал	x_i	m_i	$\frac{x_i - C}{k}$	$\frac{(x_i - C)m_i}{k}$	$\frac{(x_i - C)^2}{k^2}$	$\frac{(x_i - C)^2 m_i}{k^2}$	w_i $(m_i/120)$	$F_n^*(x)$
[7,75; 9,25)	8,5	1	–5	–5	25	25	0,0083	0,0083
[9,25; 10,75)	10,0	4	–4	–16	16	64	0,0333	0,0416
[10,75; 12,25)	11,5	12	–3	–36	9	108	0,1000	0,1416
[12,25; 13,75)	13,0	9	–2	–18	4	36	0,0750	0,2166
[13,75; 15,25)	14,5	23	–1	–23	1	23	0,1917	0,4083
[15,25; 16,75)	16,0	26	0	0	0	0	0,2167	0,6250
[16,75; 18,25)	17,5	15	1	15	1	15	0,1250	0,7500
[18,25; 19,75)	19,0	17	2	34	4	68	0,1417	0,8917
[19,75; 21,25)	20,5	12	3	36	9	108	0,1000	0,9917
[2 1,25; 22,75)	22,0	1	4	4	16	16	0,0083	1,0000
		120		–9		463	≈1,00	

График интервального вариационного ряда (относительные частоты)

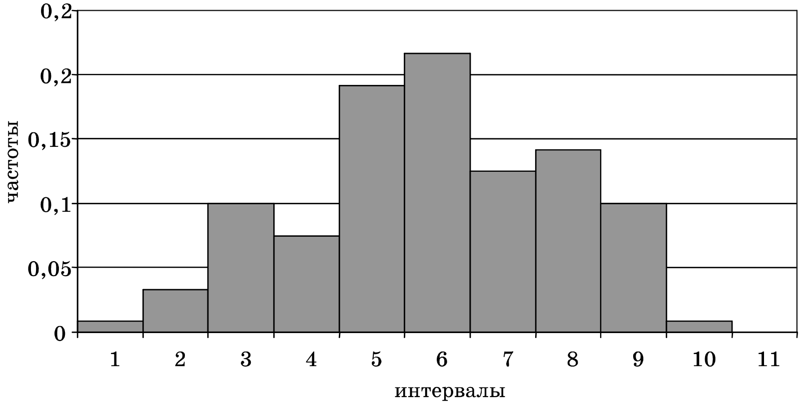
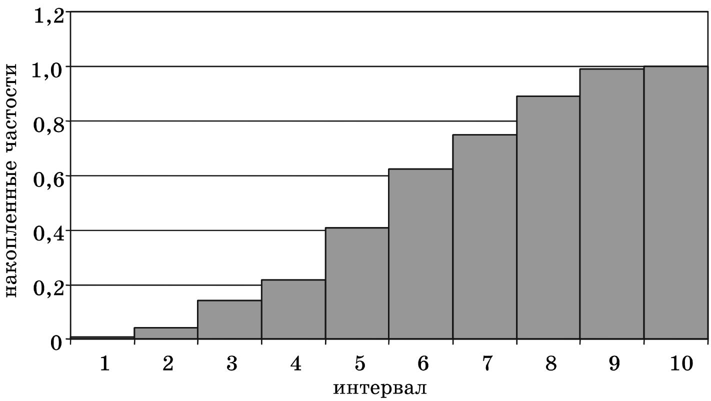


График эмпирической функции распределения



По интервальному вариационному ряду вычисляются выборочные: среднее значение, дисперсия, среднеквадратическое отклонение, мода и медиана.

Среднее значение:

$$\widehat{\widehat{X}}_n = \frac{\sum_{i=1}^{10} \frac{(x_i - C) m_i}{k}}{n} k + C$$

$$\widehat{\widehat{X}}_n = \frac{-9}{120} \times 1{,}5 + 16{,}0 = 15{,}8875 \cong 15{,}887$$

Дисперсия:

$$s_n^2 = \frac{\sum_{i=1}^{10} \frac{(x_i - C)^2}{k^2} m_i}{n} k^2 - \left(\widehat{\widehat{X}}_n - C\right)^2$$

$$s_n^2 = \frac{463}{120} \times 1{,}5^2 - (15{,}89 - 16{,}0)^2 \cong 8{,}68$$

Среднеквадратическое отклонение:

$$\sqrt{s^2} = \sqrt{\frac{n}{(n-1)} s_n^2}$$

$$s = \sqrt{\frac{120}{119} \times 8{,}68} \cong 2{,}96$$

Мода:

$$M_0 = x_0 + k \frac{m_i - m_{i-1}}{(m_i - m_{i-1}) + (m_i - m_{i+1})},$$

$m_i = 26$ — максимальная частота

$m_{i-1} = 23$

$m_{i+1} = 15$

$x_0 = 15{,}25$ — начало модового интервала

$$M_0 = 15{,}25 + 1{,}5 \frac{26 - 23}{(26 - 23) + (26 - 15)} \cong 15{,}57$$

Медиана:

$$M_e = x_0 + k\frac{\left(\frac{n}{2} - T_{i-1}\right)}{m_i},\ T_{i-1} = \sum_{j=1}^{i-1} m_j\ ,$$

$x_0 = 13{,}75$ — начало медианного интервала $\left(F_n^*(x) \le \frac{1}{2}\right)$

$T_{i-1} = 26$

$m_i = 23$ — частота медианного интервала

$$M_e = 13{,}75 + 1{,}5\frac{\left(\frac{120}{2} - 26\right)}{23} \cong 15{,}97$$

2. Используя выборки 2 и 3, по дискретному вариационному ряду вычислить несмещенные оценки для среднего значения, дисперсии, среднеквадратического отклонения генеральной совокупности.

Оценка среднего значения:

$$\widehat{\overline{X}}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Оценки дисперсии и среднеквадратического отклонения:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \widehat{\overline{X}}_n\right)^2$$

$$s = \sqrt{s^2}$$

Выборка 2:

$$\widehat{\overline{X}}_n = 14{,}858$$

$$s^2 = 7{,}94$$

$$s = 2{,}82$$

Выборка 3:

$$\hat{\hat{X}}_n = 16{,}392$$
$$s^2 = 8{,}25$$
$$s = 2{,}87$$

3. Для выборки 1, считая, что дисперсия элементов генеральной совокупности $\sigma^2 = T^2$ ($T = 4$):

– определить доверительный интервал для оценки среднего значения при доверительной вероятности $P = 1 - \alpha_1$:

$$P = 1 - \alpha_1 = 0{,}97 = 2\Phi(z) \Rightarrow \Phi(z) = 0{,}485$$
$$P = 1 - \alpha_1;\ \alpha_1 = 0{,}03$$

$$z_\alpha = \Phi^{-1}(0{,}485) = 2{,}17$$

$$\varepsilon_\alpha = \frac{z_\alpha \sigma}{\sqrt{n}};\ \varepsilon_\alpha = \frac{2{,}17 \times 4}{\sqrt{80}} = 0{,}792$$

$$\left(\hat{\hat{X}}_n - \varepsilon_\alpha;\ \hat{\hat{X}}_n + \varepsilon_\alpha\right)$$

(15,887 – 0,792; 15,887 + 0,792) (15,095; 16,679) — *доверительный интервал*

– по предельной ошибке выборки ε для оценки среднего значения найти соответствующую ему доверительную вероятность:

$$\varepsilon = 0{,}65$$

$$z_\alpha = \frac{\varepsilon\sqrt{n}}{\sigma};\ z_\alpha = \frac{0{,}65 \times \sqrt{80}}{4} \approx 1{,}78;\ \Phi(1{,}78) = 0{,}462$$

$P = 2\Phi(z) = 2\Phi(1{,}78) = 0{,}924$ — *доверительная вероятность*

– определить необходимый объем выборки для определения среднего значения генеральной совокупности с доверительной вероятностью $P = 1 - \alpha_2$ и предельной ошибкой выборки ε:

$$P = 1 - \alpha_2;\ \alpha_2 = 0{,}02$$

$$P = 2\Phi(z) = 0{,}98$$
$$\Phi(z) = 0{,}49$$
$$z_\alpha = \Phi^{-1}(z) = \Phi^{-1}(0{,}49) = 2{,}36$$

$$n = \frac{z_\alpha^2 \sigma^2}{\varepsilon^2};\ n = \frac{2{,}36^2 \times 16}{0{,}65^2} \approx 211$$ — *необходимый объем выборки*

4. Используя выборку 2, определить доверительный интервал, соответствующий доверительной вероятности $P = 1 - \alpha_3$, для оценки среднего значения генеральной совокупности.

Из задачи 2 имеем $\hat{\bar{X}}_n = 14{,}858,\ s = 2{,}82.$

$$P = 1 - \alpha_3,\ \alpha_3 = 0{,}05$$

$$t = \frac{\hat{\bar{X}}_n - \bar{X}}{s}\sqrt{n},\ P\left\{|t| < t_\alpha\right\} = 1 - \alpha_3$$

$$P\left\{\frac{\left|\hat{\bar{X}}_n - \bar{X}\right|}{s}\sqrt{n} < t_\alpha\right\} = P\left\{\hat{\bar{X}}_n - t_\alpha \frac{s}{\sqrt{n}} < \bar{X} < \hat{\bar{X}}_n + t_\alpha \frac{s}{\sqrt{n}}\right\}$$

$$S(\alpha, v)$$
$$v = n - 1 = 24$$
$$t_\alpha = 2{,}064$$
$$\varepsilon_\alpha = t_\alpha \frac{s}{\sqrt{n}} = 2{,}064 \times \frac{2{,}82}{\sqrt{25}} = 1{,}164\ ;$$
$$\hat{\bar{X}}_n - \varepsilon_a = 14{,}858 - 1{,}164 = 13{,}694$$
$$\hat{\bar{X}}_n + \varepsilon_a = 14{,}858 + 1{,}164 = 16{,}022$$

(13,694; 16,022) — *доверительный интервал*

5. Используя выборку 3, определить доверительный интервал, соответствующий доверительной вероятности $P = 1 - \alpha_4$, для оценки дисперсии генеральной совокупности.

Выборочная дисперсия из задачи 2:

$$s^2 = 8{,}25,\ n = 20$$

$$P = 1 - \alpha_4;\ \ \alpha_4 = 0{,}03$$

$$\chi^2 = \frac{s^2(n-1)}{\sigma^2}$$

$$P\left\{\chi^2 > u_2\right\} = \frac{\alpha_4}{2} = 0{,}015 \Rightarrow u_2 \approx 35,\ (v = 19;\ \alpha = 0{,}015)$$

$$P\left\{\chi^2 > u_1\right\} = 1 - \frac{\alpha_4}{2} = 0{,}985 \Rightarrow u_1 \approx 8{,}7;\ (\nu = 19,\ \alpha = 0{,}985)$$

$$P\left\{u_1 < \chi^2 < u_2\right\} = 1 - \alpha_4$$

$$\left(\frac{s^2(n-1)}{u_2};\ \frac{s^2(n-1)}{u_1}\right);\ \left(\frac{8{,}25^2(20-1)}{35};\ \frac{8{,}25^2(20-1)}{8{,}7}\right);$$

(4,479; 18,017) — *доверительный интервал*

6. Определить доверительный интервал, соответствующий доверительной вероятности P=1−α_5, для оценки доли признака. Объем выборки равен 100, выборочная доля признака равна 15+K.

n=100, K=0, P=1−α_5; α_5=0,05; m=15

$$\hat{p} = \frac{m}{n} = 0{,}15;\ P\left\{\left|p - \hat{p}\right| < \varepsilon_\alpha\right\} = 2\Phi(z) = 1 - \alpha_5,\ z_\alpha = 1{,}96$$

$$\sigma(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0{,}036;\ \varepsilon_\alpha = 1{,}96 \times 0{,}036 = 0{,}0706;$$

$$\left(\hat{p} - \varepsilon_\alpha;\ \hat{p} + \varepsilon_\alpha\right);\ (0{,}15 - 0{,}0706;\ 0{,}15 + 0{,}0706);$$

(0,0794; 0,2206) — *доверительный интервал*

7. Проверить по выборке 2 гипотезу о том, что среднее значение генеральной совокупности равно A на уровне значимости α_6 при альтернативной гипотезе — среднее значение не равно A.

$$H_0 : \bar{X} = 16$$

$$H_1 : \bar{X} \neq 16,\ n = 25,\ \hat{\hat{X}}_n = 14{,}858,\ s^2 = 7{,}94$$

$$\bar{X} = 16,\ \alpha_6 = 0{,}03,\ t = \frac{\hat{\hat{X}}_n - \bar{X}}{\sqrt{s^2}}\sqrt{n}$$

$$P\{|t| > t_\alpha\} = \alpha_6,\ P\{t < -t_\alpha\} = P\{t > t_\alpha\} = \frac{\alpha_6}{2}$$

$$t_\alpha = 2{,}349$$

$$\hat{t} = \frac{14{,}858 - 16}{2.82}\sqrt{25} = -2{,}025$$

$$\hat{t} \in (-t_\alpha; t_\alpha)$$

$$-2{,}025 \in (-2{,}349;\ 2{,}349)$$

гипотеза H_0 принимается

8. Проверить по выборке 3 гипотезу о том, что дисперсия генеральной совокупности равна T^2 на уровне значимости α_7 при альтернативной гипотезе дисперсия не равна T^2.

$$H_0 : \sigma^2 = 16$$

$$H_1 : \sigma^2 \neq 16$$

$$n = 20,\ s^2 = 8{,}25$$

$$\alpha_7 = 0{,}05$$

$$T^2 = 16$$

$$\chi^2 = \frac{(n-1)}{T^2}s^2$$

$$P\{\chi^2 > u_2\} = \frac{\alpha_7}{2} = 0{,}025 \Rightarrow u_2 = 32{,}852$$

$$P\{\chi^2 > u_1\} = 1 - \frac{\alpha_7}{2} = 0{,}075 \Rightarrow u_1 = 8{,}907$$

$$\hat{\chi}^2 = \frac{(20-1)}{16} 8{,}25 = 9{,}797$$

$$9{,}797 \in (8{,}907;\ 32{,}852)$$

гипотеза H_0 *принимается.*

9. По выборкам 2 и 3 проверить гипотезу о том, что средние значения соответствующих генеральных совокупностей равны на уровне значимости α_8 при альтернативной гипотезе — они не равны.

$$H_0 : \overline{X} = \overline{Y}$$

$$H_1 : \vec{X} \neq \overline{Y}$$

$$\alpha_8 = 0{,}02$$

$$\hat{\overline{\overline{X}}}_n = 14{,}858;\ \hat{\overline{\overline{Y}}}_n = 16{,}392$$

$$s_x^2 = 7{,}94 \quad s_y^2 = 8{,}25$$

$$\nu = n_1 + n_2 - 2 = 43$$

$$\theta = \frac{\hat{\overline{\overline{X}}}_n - \hat{\overline{\overline{Y}}}_n}{\sqrt{n_1 s_x^2 + n_2 s_y^2}} \frac{\sqrt{n_1 + n_2 - 2}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

есть t *— статистика Стьюдента*

$$t_\alpha = 2{,}418;\ \hat{\theta} = \frac{14{,}858 - 16{,}392}{\sqrt{25 \cdot 7{,}94 + 20 \cdot 8{,}25}} \frac{\sqrt{25 + 20 - 2}}{\sqrt{\frac{1}{25} + \frac{1}{20}}} = -1{,}759$$

$$\hat{\theta} = -1{,}759 \in (-2{,}418;\ 2{,}418)$$

гипотеза H_0 *принимается*

10. По выборке 1 проверить гипотезу о том, что генеральная совокупность имеет нормальное распределение с параметрами $\overline{X} = A = 16,\ \sigma^2 = T^2 = 16$ на уровне значимости α_9=0,05.

$H_0: \{\text{выборочная совокупность}\} \subset N(16;\ 16)$

$H_1: \{\text{выборочная совокупность}\} \not\subset N(16;\ 16)$

$\hat{\chi}^2 = \sum_{i=1}^{l} \frac{\left(m_{i_p} - m_i\right)^2}{m_{i_p}}$; число степеней свободы $\nu = l - r - 1$, где r — число параметров распределения; l — количество интервалов интервального вариационного ряда. $L = 10$, $r = 2$.

$P\left\{\hat{\chi}^2 > \chi^2_\alpha\right\} = 0{,}05 \Rightarrow \chi^2_\alpha = 14{,}067;\ \hat{\chi}^2 = 20{,}066 > 14{,}067.$

гипотеза H_0 не принимается.

11. С помощью метода крайних точек («натянутой нити») найти линейные функции регрессии для связанных выборок XY, XZ, YZ и построить их графики.

а) Крайние точки для выборок XY есть $M_1(9{,}27;\ 5{,}17)$ и $N_1(20{,}68;\ 10{,}86)$. Уравнение прямой через две точки на плоскости $x0y$ в канонической форме

$\frac{x - 9{,}27}{20{,}68 - 9{,}27} = \frac{y - 5{,}17}{10{,}86 - 5{,}17}$; откуда окончательно

$y = 0{,}499x + 0{,}547.$

б) Крайние точки для выборок XZ есть $M_2(9{,}27;\ 21{,}11)$ и $N_2(20{,}68;\ 10{,}85)$. Уравнение прямой через две точки на плоскости $x0z$ в канонической форме

$\frac{x - 9{,}27}{20{,}68 - 9{,}27} = \frac{z - 21{,}11}{10{,}85 - 21{,}11}$, откуда окончательно

$z = -0{,}899x + 29{,}446.$

в) Крайние точки для выборки YZ есть $M_3(5{,}17;\ 21{,}11)$ и $N_1(10{,}86;\ 10{,}85)$. Уравнение прямой на плоскости $y0z$ в канонической форме имеет вид $\frac{y - 5{,}17}{10{,}86 - 5{,}17} = \frac{z - 21{,}11}{10{,}85 - 21{,}11}$, откуда окончательно $z = -1{,}803y + 30{,}432.$

Интервал	x_i	m_i	$\frac{x_i-\bar{X}}{\sigma}$	$\Phi\left(\frac{x_i-\bar{X}}{\sigma}\right)$	$P_i=\Delta\Phi_i$	$m_{i_p}=nP_i$	$m_{i_p}-m_i$	$\left(m_{i_p}-m_i\right)^2$	$\frac{\left(m_{i_p}-m_i\right)^2}{m_{i_p}}$
(–∞; 9,25)	–∞	1	–∞	–0,5	0,0668	8,38	7,38	54,46	6,499
[9,25; 10,75)	10	4	–1,5	–0,4332	0,0635	7,62	3,62	13,10	1,719
[10,75; 12,25)	11,5	12	–1,125	–0,3697	0,0963	11,56	0,43	0,18	0,016
[12,25; 13,75)	13,0	9	–0,75	–0,2734	0,1272	15,26	6,26	39,19	2,570
[13,75; 15,25)	14,5	23	–0,375	–0,1462	0,1462	17,54	–5,46	29,81	1,700
[15,25; 16,75)	16,0	26	0	0	0,1462	17,54	–8,46	71,57	4,080
[16,75; 18,25)	17,5	15	0,375	0,1462	0,1272	15,26	0,26	0,0676	0,004
[18,25; 19,75)	19,0	17	0,750	0,2734	0,0963	11,56	–5,44	29,59	2,560
[19,75; 21,25)	20,5	12	1,125	0,3697	0,1303	15,64	3,64	13,25	0,847
[21,25; +∞)	+∞	1	+∞	0,5000	–	–	–	–	–
					1,0000				$\hat{\chi}^2=20,066$

12. По методу наименьших квадратов найти линейные функции регрессии для двумерных выборок задачи 11 и наложить их графики на графики предыдущей задачи.
Выборочные средние

$$\hat{\hat{X}}_n = \frac{1}{n}\sum_i X_i = 14{,}853;\ \hat{\hat{Y}}_n = \frac{1}{n}\sum_i Y_i = 8{,}591;\ \hat{\hat{Z}} = \frac{1}{n}\sum_i Z_{i_n} = 16{,}283.$$

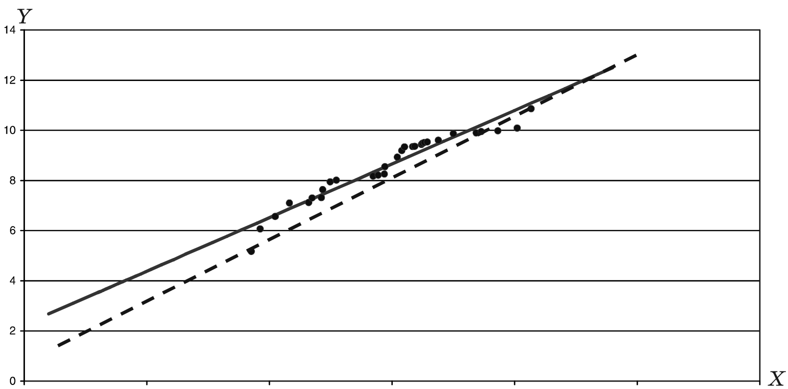
а) Для выборок XY уравнение регрессии $y = a_1 + b_1 x$,

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\hat{\hat{X}}_n \hat{\hat{Y}}_n}{\sum_{i=1}^{n} X_i^2 - n\hat{\hat{X}}_n^2} = \frac{3944{,}2 - 30\times 14{,}853\times 8{,}591}{6890{,}58 - 30\times(14{,}853)^2} \approx 0{,}427,$$

$$a_1 = \hat{\hat{Y}}_n - b_1\hat{\hat{X}}_n = 8{,}591 - 0{,}427\times 14{,}853 \approx 2{,}25.$$

Окончательно имеем $y = 0{,}427x + 2{,}25$.

Графики парной регрессии Y на X



– – по методу крайних точек
—— по методу наименьших квадратов

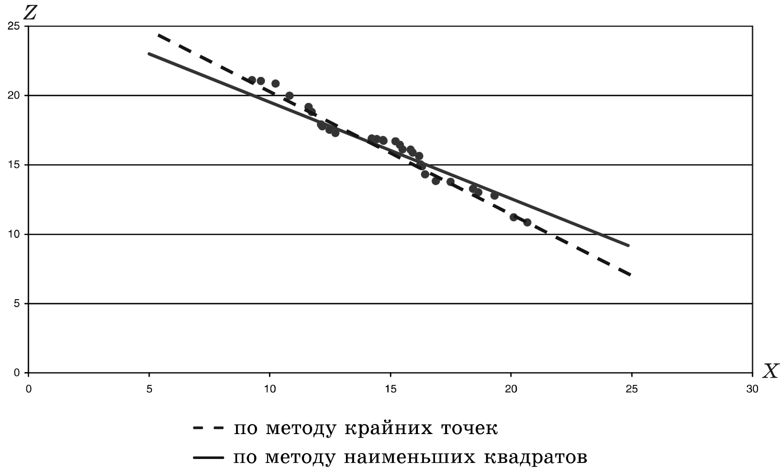
б) Для выборок XZ уравнение регрессии $z = a_2 + b_2 x$,

$$b_2 = \frac{\sum_{i=1}^{n} X_i Z_i - n\bar{X}_n \bar{Z}_n}{\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2} = \frac{7019{,}03 - 30 \times 14{,}853 \times 16{,}283}{6890{,}58 - 30 \times (14{,}853)^2} \approx -0{,}87,$$

$$a_2 = \bar{Z}_n - b_2 \bar{X}_n = 16{,}283 - (-0{,}87) \times 14{,}853 \approx 29{,}201.$$

Окончательно имеем $z = -0{,}87x + 29{,}201$.

Графики парной регрессии Z на X



в) Для выборок YZ уравнение регрессии $z = a_3 + b_3 y$.

$$b_3 = \frac{\sum_{i=1}^{n} Y_i Z_i - n\bar{Y}_n \bar{Z}_n}{\sum_{i=1}^{n} Y_i^2 - n\bar{Y}_n^2} = \frac{4090{,}43 - 30 \times 8{,}591 \times 16{,}283}{2267{,}44 - 30 \times (8{,}591)^2} \approx -2{,}00.$$

$$a_3 = \bar{Z}_n - b_3 \bar{Y}_n = 16{,}283 - (-2{,}00)8{,}591 = 33{,}465.$$

Окончательно имеем $z = -2y + 33{,}465$.

13. Полагая, что каждый i-й столбец в выборке 2 соответствует i-му уровню фактора A_i, оценить влияние фактора A на уровне значимости α_{10} при полностью случайном плане эксперимента.

Выборку 2 представляем как наблюдаемые значения признака.

Номер уровня j фактора	Наблюдаемые значения признака					Объем выборки n_j	Сумма T_j	Групповая средняя $\hat{\bar{X}}_j$
1	9,27	15,51	15,84	18,60	12,73	5	71,95	14,39
2	16,32	14,72	11,75	16,25	14,69	5	73,73	14,75
3	15,22	20,11	8,43	12,48	14,44	5	80,68	16,14
4	16,89	19,32	9,63	15,93	12,12	5	73,89	14,78
p=5	14,23	10,82	17,50	16,44	12,18	5	71,17	14,23
	Итого					25	371,42	$\hat{\bar{\bar{X}}}=14,86$

$$Q_0=\sum_i\sum_j\left(X_{ij}-\hat{\bar{\bar{X}}}\right)^2=204,6;$$

$$Q_2=\sum_j n_j\left(\hat{\bar{X}}_g-\hat{\bar{\bar{X}}}\right)^2=11,36;$$

$$Q_1=Q_0-Q_2=204,6-11,36=193,24.$$

$$\nu_0=n-1=25-1=24;\ \nu_2=p-1=5-1=4;$$

$$\nu_1=n-p=25-5=20.$$

$$s_0^2=\frac{Q}{\nu_0}=\frac{204,6}{24}\approx 8,525;\ s_1^2=\frac{Q_1}{\nu_1}=\frac{193,24}{20}=9,66;$$

$$s_2^2=\frac{Q_2}{\nu_2}=\frac{11,36}{4}=2,84.$$

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2,\ \alpha_{10} = 0{,}05;\ F = \frac{s_2^2}{s_1^2};$$

$$F_\alpha = 2{,}87;\ \widehat{F} = \frac{2{,}84}{9{,}66} \approx 0{,}29;\ 0{,}29 < 2{,}87,$$

Тогда H_0 *принимается, т. е. фактор А не влияет.*

14. Произвести сглаживание заданного временного ряда методом скользящих средних, выявить тренд с использованием метода наименьших квадратов и найти прогноз на 3 временные единицы. Построить графики.

Принимаем величину скользящего временного интервала $m = 3$, тогда сглаженные значения определяются по формуле

$$\hat{y}_{t_{cc}} = \frac{1}{3}\sum_{i=t-1}^{t+1} y_i,$$

Время	1	2	3	4	5	6	7	8	9	10	11	12
Параметр	–	36,9	36,3	38,3	39,3	40,0	40,7	41,3	42,7	44,1	46,9	46,0

Время	13	14	15	16	17	18	19	20	21	22	23	24
Параметр	48,7	49,2	49,3	48,7	50,9	54,7	55,6	53,4	53,8	55,7	58,9	–

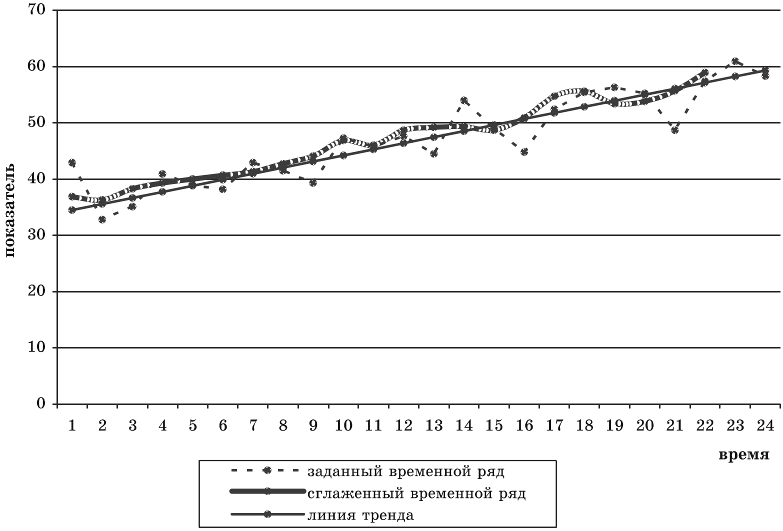
Принимаем гипотезу о линейном характере тренда. Тогда модель $y = a + bt$,

$$\hat{\hat{T}}_n = \frac{1}{n}\sum_{i=1}^{n} t_i = \frac{1}{24}295 \approx 12{,}29,\ \hat{\hat{Y}}_n = \frac{1}{n}\sum y_i = \frac{1}{24}1129{,}92 = 47{,}08,$$

$$b = \frac{\sum_{i=1}^{n} T_i Y_i - n\hat{\hat{T}}_n \hat{\hat{Y}}_n}{\sum_{i=1}^{n} T_I^2 - n\hat{\hat{T}}^2} = \frac{15268{,}5 - 24 \times 12{,}29 \times 47{,}08}{4900{,}08 - 24 \times (12{,}29)^2} \approx 1{,}08,$$

$$a = \hat{\hat{Y}}_n - b\hat{\hat{T}}_n = 47{,}08 - 1{,}08 \times 12{,}29 \approx 34{,}5.$$

Уравнение тренда: $y(t) = 34{,}5 + 1{,}08t$.



Прогноз на 3 временные единицы: $y_{27} = 34{,}5 + 1{,}08 \times 27 = 63{,}66$.

Приложение 2

Функция Лапласа $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_0^z e^{-t^2/2}\,dt$, где $z=\frac{x-a}{\sigma}$

z	Сотые доли для z									
	0	1	2	3	4	5	6	7	8	9
00	0,0000	,0040	,0080	,0120	,0160	,0199	,0239	,0279	,0319	,0359
0,1	,0398	,0438	,0478	,0517	,0557	,0596	,0636	,0675	,0714	,0753
0,2	,0793	,0832	,0871	,0910	,0948	,0987	,1026	,1064	,1103	,1141
0,3	0,1179	,1217	,1255	,1293	,1331	,1368	,1406	,1443	,1480	,1517
0,4	,1554	,1591	,1628	,1664	,1700	,1736	,1772	,1808	,1844	,1879
0,5	,1915	,1950	,1985	,2019	,2054	,2088	,2123	,2157	,2190	,2224
0,6	0,2257	,2291	,2324	,2357	,2389	,2422	,2454	,2486	,2517	,2549
0,7	,2580	,2611	,2642	,2673	,2703	,2734	,2764	,2794	,2823	,2852
0,8	,2881	,2910	,2939	,2967	,2995	,3023	,3051	,3078	,3106	,3133
0,9	0,3159	,3186	,3212	,3238	,3264	,3289	,3315	,3340	,3365	,3389
1,0	,3413	,3437	,3461	,3485	,3508	,3531	,3554	,3577	,3599	,3621
1,1	,3643	,3665	,3686	,3708	,3729	,3749	,3770	,3790	,3810	,3830
1,2	,3849	,3869	,3888	,3907	,3925	,3944	,3962	,3980	,3997	,4015
1,3	0,4032	,4049	,4066	,4082	,4099	,4115	,4131	,4147	,4162	,4177
1,4	,4192	,4207	,4222	,4236	,4251	,4265	,4279	,4292	,4306	,4319
1,5	,4332	,4345	,4357	,4370	,4382	,4394	,4406	,4418	,4429	,4441
1,6	,4452	,4463	,4474	,4484	,4495	,4505	,4515	,4525	,4535	,4545
1,7	,4554	,4564	,4573	,4582	,4591	,4599	,4608	,4616	,4625	,4633
1,8	,4641	,4649	,4656	,4664	,4671	,4678	,4686	,4693	,4699	,4706
1,9	,4713	,4719	,4726	,4732	,4738	,4744	,4750	,4756	,4761	,4767
2,0	,4772	,4778	,4783	,4788	,4793	,4798	,4803	,4808	,4812	,4817
2,1	,4821	,4826	,4830	,4834	,4838	,4842	,4846	,4850	,4854	,4857
2,2	,4860	,4864	,4867	,4871	,4874	,4877	,4880	,4883	,4886	,4889
2,3	,4892	,4895	,4898	,4900	,4903	,4906	,4908	,4911	,4913	,4915
2,4	,4918	,4920	,4922	,4924	,4926	,4928	,4930	,4932	,4934	,4936
2,5	,4937	,4939	,4941	,4942	,4944	,4946	,4947	,4949	,4950	,4952
2,6	,4953	,4954	,4956	,4957	,4958	,4959	,4960	,4962	,4963	,4964
2,7	,4965	,4966	,4967	,4968	,4969	,4970	,4971	,4971	,4972	,4973
2,8	0,4974	,4975	,4975	,4976	,4977	,4978	,4978	,4979	,4980	,4980
2,9	,4981	,4981	,4982	,4983	,4983	,4984	,4984	,4985	,4985	,4986
3,0	,4986	,4986	,4987	,4987	,4988	,4988	,4988	,4989	,4989	,4989
3,1	,4990	,4990	,4990	,4991	,4991	,4991	,4992	,4992	,4992	,4992
3,2	,4993	,4993	,4993	,4993	,4994	,4994	,4994	,4994	,4994	,4994
3,3	,4995	,4995	,4995	,4995	,4995	,4995	,4996	,4996	,4996	,4996
3,4	,4996	,4996	,4996	,4996	,4997	,4997	,4997	,4997	,4997	,4997
3,5	,4997	,4997	,4997	,4997	,4997	,4998	,4998	,4998	,4998	,4998
3,6	,4998	,4998	,4998	,4998	,4998	,4998	,4998	,4998	,4998	,4998
3,7	,4998	,4998	,4999	,4999	,4999	,4999	,4999	,4999	,4999	,4999

Приложение 3

Пределы для χ^2-распределения Пирсона

ν \ α	**0,995**	**0,975**	**0,95**	**0,05**	**0,025**	**0,005**
1	$0{,}39\cdot10^{-4}$	$0{,}98\cdot10^{-3}$	$0{,}39\cdot10^{-2}$	3,841	5,024	7,879
2	0,010	0,050	0,103	5,991	7,378	10,597
3	0,071	0,216	0,352	7,815	9,348	12,838
4	0,207	0,484	0,711	9,488	11,143	14,830
5	0,412	0,831	1,145	11,070	12,832	16,750
6	0,676	1,237	1,635	12,592	14,449	18,475
7	0,989	1,690	2,167	14,067	16,013	20,278
8	1,314	2,180	2,733	15,507	17,535	21,955
9	1,735	2,700	3,325	16,919	19,023	23,589
10	2,156	3,247	3,940	18,307	20,483	25,188
11	2,603	3,816	4,575	19,675	21,920	26,757
12	3,074	4,404	5,226	21,026	23,336	28,300
13	3,565	5,009	5,892	22,362	24,736	29,819
14	4,075	5,629	6,571	23,685	26,119	31,319
15	4,601	6,262	7,261	24,996	27,488	32,804
16	5,142	6,908	7,962	26,296	28,845	34,267
17	5,697	7,564	8,672	27,587	30,191	35,713
18	6,265	8,231	9,390	28,869	31,526	37,156
19	6,844	8,907	10,117	30,144	32,852	38,582
20	7,434	9,591	10,851	31,410	34,170	39,897
21	8,034	10,283	11,591	32,671	35,479	41,401
22	8,643	10,982	12,338	33,924	36,781	42,796
23	9,260	11,688	13,091	35,172	38,076	44,181
24	9,886	12,401	13,848	36,415	39,364	45,558
25	10,520	13,120	14,671	37,652	40,646	46,928
26	11,760	13,844	15,379	38,885	41,923	48,290
27	11,808	14,573	16,151	40,113	42,194	49,645
28	12,461	14,308	16,928	41,337	44,461	50,993
29	13,121	16,047	17,708	42,557	45,722	52,336
30	13,787	16,791	18,498	43,773	46,979	53,672
31	14,458	17,539	19,281	44,985	48,232	53,003
32	15,134	18,291	20,072	46,194	49,483	56,328
33	15,815	19,047	20,867	47,400	50,725	57,648
34	16,501	19,803	21,664	48,602	51,966	58,964
35	17,192	20,569	22,465	49,802	53,203	60,275
36	17,887	21,336	23,269	50,998	54,437	61,581
37	18,586	22,106	24,075	52,192	55,668	62,882
38	19,289	22,878	24,884	53,384	56,895	64,181
39	19,996	23,654	25,695	54,572	58,120	65,476
40	20,707	24,433	26,509	55,758	59,342	66,766

Приложение 4

Двухсторонние пределы для t-распределения Стьюдента

$$P\left\{|t|>|t_{\text{кр}}|\right\}=\alpha$$

ν \ α	**0,20**	**0,10**	**0,05**	**0,02**	**0,01**	**0,005**	**0,002**	**0,001**
1	3,0770	6,3130	12,7060	31,8200	63,6560	127,6560	318,3080	636,6190
2	1,8850	2,9200	4,3020	6,9640	9,9240	14,0890	22,3270	31,5990
3	6377	3534	3,1820	4,5400	5,8400	7,4580	10,2140	12,9240
4	5332	1318	2,7760	3,7460	4,6040	5,5970	7,1730	8,6100
5	4759	0150	5706	3649	0321	4,7730	5,8930	6,8690
6	1,4390	1,9430	2,4460	3,1420	3,7070	4,3160	5,2070	5,9580
7	4149	8946	3646	2,9980	4995	0293	4,7850	4079
8	3998	8595	3060	8965	3554	3,8320	5008	0413
9	3830	8331	2622	8214	2498	6897	2968	4,7800
10	3720	8125	2281	7638	1693	5814	1437	5869
11	1,3630	1,7960	2,2010	2,7180	3,1050	3,4980	4,0240	4,4370
12	3562	7823	1788	6810	0545	4284	3,929	3178
13	3502	7709	1604	6503	0123	3725	8520	2208
14	3450	7613	1448	6245	2,976	3257	7874	1405
15	3406	7530	1314	6025	9467	2860	7328	0728
16	1,3360	1,7460	2,119	2,5830	2,9200	3,2520	3,6860	4,0150
17	3334	7396	1098	5668	8982	2224	6458	3,9650
18	3304	7341	1009	5514	8784	1966	6105	9216
19	3277	7291	0930	5395	8609	1737	5794	8834
20	3253	7247	0860	5280	8453	1534	5518	8495
21	1,3230	1,7210	2,079	2,5170	2,8310	8,1350	3,5270	8,8190
22	3212	7167	0739	5083	8188	1188	5050	7921
23	3195	7139	0687	4999	8073	1040	4850	7676
24	3178	7109	0639	4922	7969	0905	4668	7454
25	3163	7081	0595	4851	7874	0782	4502	7251
26	1,3150	1,7060	2,055	2,4780	2,7780	3,0660	3,4350	3,7060
27	3137	7033	0518	4727	7707	0565	4210	6896
28	3125	7011	0484	4671	7633	0469	4082	6739
29	3114	6991	0452	4620	7564	0380	3962	6594
30	3104	6973	0423	4573	7500	0298	3852	6466
32	1,3080	1,6930	2,0360	2,4480	2,7380	3,0140	3,3650	3,6210
34	3070	6909	0322	4411	7284	0020	3479	6007
36	3050	6883	0281	4345	7195	2,9900	3326	5821
38	3042	6860	0244	4286	7116	9808	3190	5657
40	3031	6839	0211	4233	7045	9712	3069	5510
42	1,3200	1,6820	2,0180	2,4180	2,6980	2,6930	3,2960	3,5370
44	3011	6802	0154	4141	6923	9555	2861	5258
46	3002	6787	0129	4102	6870	9488	2771	5150

ν \ α	**0,20**	**0,10**	**0,05**	**0,02**	**0,01**	**0,005**	**0,002**	**0,001**
48	2994	6772	0106	4066	6822	9426	2689	5051
50	2987	6759	0086	4033	6778	9370	2614	4960
55	1,9970	1,6730	2,0040	2,3960	2,6680	2,9240	3,2560	3,4760
60	2958	6706	0003	3901	6603	9146	2317	4602
65	2947	6686	1,9970	3851	6536	9060	2204	4466
70	2938	6669	9944	3808	6479	8987	2108	4350
80	1,2920	1,6640	1,9900	2,3730	2,6380	2,8870	3,1950	3,4160
90	2910	6620	9867	3885	6316	8779	1833	4019
100	2901	6602	9840	3642	6259	8707	1737	3905
120	2886	6577	9799	3578	6174	8599	1595	3735
150	2872	6551	9759	3515	6090	8492	1456	3586
200	2858	6525	9719	3451	6006	8385	1315	3398
250	2849	6510	9695	3414	5956	8222	1232	3299
300	2844	6499	9679	3388	5923	8279	1176	3233
400	2837	6487	9659	3357	5882	8227	1107	3150
500	1,2830	1,6470	1,9640	2,333	2,7850	2,8190	3,1080	3,3100

Приложение 5

Верхние односторонние пределы $F_{кр}$ для распределения Фишера-Снедекора

а) Уровень значимости $\alpha = 0,05$

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	16	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	253	253	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,41	19,43	19,44	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,74	8,69	8,66	8,64	8,62	8,60	8,57	8,56	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,84	5,80	5,77	5,74	5,71	5,68	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,47	4,68	4,60	4,56	4,53	4,50	4,46	4,42	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,92	3,87	3,84	3,81	3,77	3,72	3,71	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,679	3,73	8,68	3,63	3,57	3,49	3,44	3,41	3,38	3,34	3,279	3,28	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,28	3,20	3,15	3,12	3,08	3,05	3,00	2,98	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,07	2,98	2,93	2,90	2,86	2,82	2,77	2,76	2,71
10	4,86	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,91	2,82	2,77	2,74	2,70	2,67	2,61	2,59	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,79	2,70	2,65	2,61	2,57	2,53	2,47	2,45	2,40
12	4,75	3,88	3,49	3,26	3,14	3,00	2,92	2,85	2,80	2,76	2,69	2,60	2,54	2,50	2,46	2,42	2,36	2,35	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,60	2,51	2,46	2,42	2,38	2,34	2,28	2,26	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,53	2,44	2,39	2,35	2,31	2,27	2,21	2,19	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,48	2,39	2,33	2,29	2,25	2,21	2,15	2,12	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,33	2,28	2,24	2,20	2,16	2,09	2,07	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,38	2,29	2,23	2,19	2,15	2,11	2,04	2,02	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,25	2,19	2,15	2,11	2,07	2,00	1,98	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,31	2,21	2,15	2,11	2,07	2,02	1,96	1,94	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,28	2,18	2,12	2,08	2,04	1,89	1,92	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,15	2,09	2,05	2,00	1,96	1,89	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,20	2,23	2,13	2,07	2,03	1,98	1,93	1,87	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,20	2,10	2,04	2,00	1,96	1,91	1,84	1,82	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,18	2,09	2,02	1,98	1,94	1,89	1,82	1,80	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,28	2,24	2,16	2,06	2,00	1,96	1,92	1,87	1,80	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,05	1,99	1,90	1,95	1,85	1,78	1,76	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,20	2,13	2,03	1,97	1,93	1,88	1,84	1,76	1,74	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,12	2,02	1,96	1,91	1,87	1,81	1,75	1,72	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,18	2,10	2,00	1,94	1,90	1,85	1,80	1,73	1,71	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,09	1,99	1,93	1,89	1,84	1,79	1,72	1,69	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,00	1,90	1,84	1,79	1,74	1,69	1,61	1,59	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,81	1,75	1,70	1,65	1,59	1,50	1,48	1,39
120	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,95	1,90	1,83	1,72	1,65	1,60	1,65	1,55	1,49	1,36	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,83	1,75	1,64	1,57	1,52	1,46	1,39	1,28	1,24	1,00

б) Уровень значимости $\alpha = 0{,}025$

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	16	20	24	30	40	60	120	∞
1	647,7	799,5	864,1	899,5	921,8	937,1	948,2	956,6	963,2	968,6	976,6	984,8	993,1	997,2	1001,4	1005,6	1009,8	1014,0	1018,3
2	38,50	39,00	39,16	39,24	39,29	39,33	39,35	39,37	38,38	39,39	39,41	39,43	39,44	39,45	39,46	37,47	39,48	39,49	39,49
3	17,44	16,04	15,43	15,10	14,88	14,73	14,62	14,57	14,47	14,41	14,33	14,25	14,16	14,12	14,08	14,08	13,99	13,94	13,90
4	12,71	10,64	9,979	9,604	9,364	9,197	9,074	8,979	8,904	8,843	8,751	8,656	8,559	8,510	8,461	8,411	8,360	8,309	8,257
5	10,00	8,433	7,763	7,387	7,146	6,977	6,853	6,757	6,681	6,619	6,524	6,422	6,328	6,278	6,226	6,175	6,122	6,069	6,015
6	8,813	7,259	6,598	6,227	6,987	5,819	5,695	5,599	5,523	5,461	5,366	5,268	5,168	5,117	5,065	5,012	4,958	4,904	4,849
7	8,072	6,541	5,889	5,522	5,285	5,118	4,994	4,899	4,823	4,761	4,665	4,567	4,466	4,415	4,362	4,308	4,254	4,198	4,142
8	7,570	6,059	5,416	5,052	4,817	4,651	4,528	4,433	4,357	4,295	4,199	4,101	3,999	3,947	3,894	3,839	3,784	3,727	3,670
9	7,209	5,714	5,078	4,718	4,184	4,319	4,197	4,102	4,026	3,963	3,868	3,769	3,666	3,614	3,560	3,505	3,449	3,391	3,332
10	6,936	5,456	4,825	4,468	4,236	4,072	3,949	3,854	3,779	3,716	3,620	3,521	3,418	3,365	3,311	3,255	3,198	3,139	3,079
11	6,724	5,255	4,630	4,275	4,044	3,880	3,758	3,663	3,587	3,525	3,429	3,329	3,226	3,172	3,117	3,061	3,003	2,944	2,882
12	6,553	5,095	4,474	4,121	3,891	3,728	3,606	3,511	3,435	3,373	3,277	3,177	3,072	3,018	2,963	2,906	2,847	2,787	2,724
13	6,414	4,965	4,347	3,995	3,766	3,614	3,482	3,388	3,312	3,249	3,153	3,052	2,947	2,893	2,837	2,779	2,720	2,659	2,595
14	6,297	4,856	4,241	3,891	3,663	3,501	3,379	3,285	3,209	3,146	3,050	2,949	2,843	2,788	2,732	2,674	2,614	2,551	2,487
15	6,199	4,765	4,152	3,804	3,576	3,414	3,293	3,198	3,122	3,050	2,963	2,862	2,755	2,700	2,643	2,585	2,524	2,461	2,395
16	6,155	4,686	4,976	3,729	3,502	3,340	3,219	3,124	3,048	2,986	2,889	2,787	2,680	2,625	2,567	2,508	2,447	2,383	2,316
17	6,042	4,618	4,012	3,664	3,437	3,276	3,155	3,061	2,984	2,922	2,824	2,723	2,615	2,559	2,502	2,442	2,380	2,315	2,247
18	5,978	4,559	3,953	3,608	3,382	3,220	3,090	3,005	2,929	2,866	2,768	2,666	2,559	2,502	2,444	2,384	2,321	2,255	2,185
19	5,921	4,507	3,903	3,558	3,332	3,171	3,050	2,956	2,880	2,817	2,719	2,617	2,508	2,452	2,383	2,332	2,269	2,203	2,133
20	5,871	4,461	3,858	3,514	3,289	3,128	3,007	3,912	2,836	2,773	2,675	2,573	2,464	2,407	2,348	2,287	2,223	2,156	2,085
21	5,826	4,419	3,818	3,474	3,250	3,089	2,968	2,974	2,797	2,734	2,636	2,533	2,424	2,367	2,306	2,246	2,181	2,114	2,042
22	5,786	3,382	3,782	3,440	3,125	3,054	2,933	2,889	2,762	2,699	2,601	2,498	2,389	2,331	2,271	2,209	2,144	2,076	2,033
23	5,749	3,349	3,750	3,408	3,183	3,023	2,902	2,807	2,731	2,668	2,569	2,466	2,356	2,298	2,238	2,176	2,110	2,041	1,967
24	5,716	4,318	3,721	3,379	3,154	2,994	2,873	2,779	2,702	2,639	2,541	2,437	2,327	2,269	2,209	2,146	2,079	2,009	1,935
25	5,686	4,290	3,694	3,353	3,128	2,968	2,847	2,753	2,676	2,613	2,514	2,411	2,300	2,242	2,181	2,118	2,051	1,981	1,905
26	5,658	4,265	3,669	3,328	3,104	2,944	2,824	2,724	2,649	2,589	2,490	2,386	2,275	2,217	2,156	2,092	2,025	1,954	1,878
27	5,633	4,242	3,647	3,306	3,082	2,922	2,802	2,707	2,630	2,567	2,468	2,344	2,253	2,194	2,133	2,069	2,001	1,929	1,852
28	5,609	4,220	3,626	3,286	3,062	2,902	2,782	2,687	2,610	2,547	2,446	2,343	2,232	2,173	2,112	2,047	1,979	1,907	1,829
29	5,587	4,200	3,607	3,267	3,043	2,884	2,763	2,668	2,591	2,528	2,429	2,324	2,213	2,154	2,092	2,027	1,959	1,886	1,807
30	5,567	4,182	3,589	2,249	3,026	2,866	2,746	2,651	2,574	2,511	2,412	2,307	2,195	2,135	2,073	2,009	1,940	1,866	1,786
40	5,423	4,051	3,463	3,126	2,903	2,744	2,623	2,528	2,451	2,388	2,288	2,181	2,067	2,006	1,942	1,875	1,802	1,724	1,737
60	5,285	3,925	3,342	3,007	2,786	2,627	2,506	2,411	4,334	2,270	2,160	2,061	1,944	1,881	1,815	1,744	1,666	1,581	1,482
120	5,152	3,804	3,227	2,894	2,515	2,524	2,299	2,299	2,221	2,157	2,054	1,945	1,824	1,759	1,689	1,614	1,529	1,432	1,310
∞	5,023	3,688	3,116	2,786	2,566	2,408	2,237	2,191	2,113	2,048	1,944	1,832	1,708	1,640	1,566	1,483	1,388	1,268	1,000

в) Уровень значимости $\alpha = 0{,}01$

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	16	20	24	30	40	60	120	∞
1	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6106	6159	6208	6234	6258	6286	6323	6334	6366
2	98,49	99,05	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,42	99,44	99,45	99,47	99,47	99,48	99,49	99,49	99,50
3	23,12	30,81	29,46	28,71	28,24	27,91	26,67	27,49	27,34	27,23	27,05	26,83	26,69	26,60	26,50	26,41	26,27	26,23	26,12
4	21,20	18,00	16,69	15,28	15,52	15,21	14,90	14,80	14,66	14,54	14,37	14,15	14,02	13,93	13,83	13,74	13,61	13,57	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,89	9,68	9,55	9,47	9,38	9,29	9,17	9,13	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,52	7,52	7,39	7,31	7,23	7,14	7,02	6,69	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,47	6,27	6,15	6,07	5,98	5,90	5,78	5,75	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,67	5,48	5,36	5,28	5,20	5,11	5,00	4,96	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,11	4,92	4,80	4,73	4,64	4,66	4,45	4,41	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,71	4,52	4,41	4,33	4,25	4,17	4,05	4,01	3,91
11	9,85	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,40	4,21	4,10	4,02	3,94	3,86	3,74	3,70	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,16	4,98	3,86	3,70	3,70	3,61	3,49	3,46	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,78	3,67	3,59	3,51	3,42	3,30	3,27	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,62	3,51	3,43	3,34	3,26	3,14	3,11	3,00
15	8,68	6,38	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,48	3,36	3,29	3,20	3,12	3,00	2,97	2,87
16	8,53	6,23	5,29	4,77	4,44	4,29	4,03	3,89	3,78	3,69	3,55	3,37	3,25	3,18	3,10	3,01	2,89	2,86	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,45	3,27	3,16	3,08	3,00	2,92	2,79	2,76	2,65
18	8,28	6,01	5,09	4,58	4,23	4,01	3,85	3,71	3,60	3,51	3,37	3,19	3,07	3,00	2,91	2,83	2,71	2,68	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,20	3,12	3,00	2,92	2,84	2,76	2,63	2,60	2,49
20	8,10	5,85	4,94	4,43	4,10	8,87	3,71	3,56	3,45	3,37	3,23	3,05	2,94	2,86	2,77	2,68	2,56	2,53	2,42
21	8,02	5,78	4,87	4,37	4,04	8,81	3,65	3,51	3,40	3,31	3,17	2,99	2,88	2,80	2,72	2,63	2,51	2,47	2,36
22	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,94	2,83	2,45	2,67	2,58	2,46	2,42	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,22	3,07	2,89	2,78	2,70	2,62	2,53	2,41	2,37	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17	3,03	2,85	2,74	2,66	2,58	2,49	2,36	2,33	2,21
25	7,77	5,57	4,68	4,18	8,86	3,63	3,46	3,32	3,21	3,13	2,99	2,81	2,70	2,62	2,54	2,45	2,32	2,29	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,17	3,09	2,96	2,77	2,66	2,58	2,50	2,41	2,28	2,25	2,13
27	7,68	5,49	4,60	4,11	3,79	3,56	3,39	3,26	3,14	3,06	2,93	2,74	2,63	2,55	2,47	2,38	2,25	2,21	2,10
28	7,64	5,45	4,57	4,07	3,76	3,53	3,36	8,23	3,11	3,03	2,90	2,71	2,60	2,52	2,44	2,35	2,22	2,18	2,06
29	7,60	5,42	4,54	4,04	3,73	8,50	3,33	3,20	3,08	3,00	2,87	2,68	2,57	2,49	2,41	2,32	2,19	2,15	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,84	2,66	2,55	2,47	2,38	2,29	2,16	2,13	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,88	2,80	2,66	2,49	2,37	2,29	2,20	2,11	1,91	1,94	1,81
60	7,08	4,08	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,67	2,50	2,32	2,20	2,12	2,03	1,93	1,97	1,74	1,60
120	6,84	4,78	3,94	3,47	3,17	2,95	2,79	2,65	2,56	2,47	2,33	2,15	2,03	1,94	1,85	1,75	1,59	1,54	1,37
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	1,99	1,87	1,79	1,69	1,59	1,41	1,36	1,09

г) Уровень значимости $\alpha = 0{,}005$

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	16	20	24	30	40	60	120	∞
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	198,5	199,0	199,1	199,2	199,3	199,3	199,3	199,3	199,3	199,4	199,4	199,4	199,4	199,4	199,4	199,4	199,4	199,4	199,4
3	55,55	49,79	47,46	46,19	45,89	44,83	41,43	44,12	43,88	43,68	43,38	43,09	42,77	42,62	42,46	42,30	42,14	41,98	41,82
4	81,33	26,28	24,25	23,15	22,45	21,97	21,62	21,35	21,13	20,96	20,70	20,43	20,16	20,03	19,89	19,75	19,61	19,46	19,32
5	22,78	18,31	16,53	15,55	14,91	14,51	13,20	13,95	13,77	13,61	13,38	13,14	12,90	12,78	12,65	12,53	12,40	12,27	12,14
6	18,63	15,54	12,91	12,07	11,46	11,07	10,78	10,56	10,39	10,25	10,08	9,814	9,588	9,474	9,358	9,240	2,121	2,001	8,879
7	16,23	12,40	10,88	10,05	9,522	9,155	8,885	8,678	8,513	8,380	8,176	7,967	7,754	7,645	7,534	7,422	7,308	7,193	7,076
8	14,68	11,04	9,596	8,805	8,301	7,952	7,961	7,496	7,338	7,210	7,014	6,814	6,608	6,502	6,396	6,287	6,177	8,064	5,950
9	13,61	10,10	8,717	7,955	7,471	7,133	6,884	6,693	6,541	6,417	6,227	6,032	5,831	5,729	5,524	5,518	5,410	5,300	5,187
10	12,82	9,427	8,080	7,342	6,872	6,514	6,302	6,115	5,967	5,846	5,661	5,470	5,274	5,173	5,070	4,965	4,859	4,750	4,638
11	12,22	8,912	7,600	6,880	6,421	6,101	5,864	5,682	5,536	5,418	5,236	5,048	4,855	4,755	4,654	4,550	4,445	4,336	4,225
12	11,75	8,509	7,225	6,521	6,071	5,757	5,524	5,345	5,202	5,085	4,906	4,721	4,529	4,431	4,330	4,228	4,122	4,014	3,903
13	11,37	8,186	6,925	6,233	5,791	5,481	5,252	5,076	4,935	4,819	4,642	4,460	4,270	4,172	3,970	3,865	3,757	3,757	3,646
14	11,06	7,921	6,680	5,998	5,562	5,257	5,031	4,856	4,717	4,603	4,428	4,246	4,058	3,961	3,861	3,760	3,655	3,547	3,435
15	10,79	7,700	6,476	6,802	5,372	5,078	5,847	4,674	4,536	4,423	4,249	4,069	3,882	3,785	3,686	3,585	3,460	3,372	3,260
16	10,57	7,513	6,303	5,537	5,211	4,913	4,092	4,520	4,383	4,271	4,099	3,960	3,734	3,637	3,538	3,437	3,332	3,224	3,111
17	10,38	7,353	6,155	5,496	5,074	4,778	4,559	4,389	4,253	4,142	3,970	3,792	3,607	3,511	3,412	3,310	3,205	3,097	2,283
18	10,21	7,214	6,027	5,374	4,956	4,662	4,441	4,275	4,141	4,030	3,859	3,683	3,497	3,401	3,303	2,201	3,096	2,987	2,873
19	10,07	7,095	5,916	5,268	4,852	4,561	4,344	4,177	4,042	3,932	3,763	3,586	3,402	3,306	3,207	3,405	3,000	2,890	2,776
20	9,943	6,986	5,817	5,174	4,761	4,472	4,256	4,090	3,956	3,847	3,677	3,502	3,317	3,222	3,123	3,021	2,915	2,805	2,690
21	9,829	6,891	5,730	5,091	4,680	4,393	4,178	4,012	3,878	3,770	3,602	3,427	3,243	3,147	3,048	2,946	2,640	2,730	2,614
22	9,727	6,806	5,652	5,016	4,608	4,322	4,109	3,944	3,811	3,703	3,535	3,360	3,176	3,080	2,982	2,879	2,773	2,661	2,545
23	9,634	6,730	5,582	4,950	4,544	4,259	4,046	3,882	3,750	3,642	3,474	2,299	3,116	3,020	2,922	2,819	2,713	2,601	2,483
24	9,551	6,661	5,519	4,889	4,485	4,201	3,990	3,826	3,694	3,587	3,419	3,245	3,062	2,966	2,867	2,765	2,658	2,546	2,427
25	9,475	6,598	5,461	4,836	4,432	4,150	3,939	3,755	3,644	3,537	3,370	3,196	3,913	2,917	2,818	2,716	2,608	2,496	2,376
26	9,405	6,540	5,409	4,785	4,384	4,102	3,892	3,729	3,598	3,491	3,325	3,151	2,968	2,872	2,773	2,670	2,563	2,450	2,329
27	9,342	6,488	5,361	4,739	4,340	4,059	3,350	3,685	3,557	3,440	3,283	3,110	2,927	2,831	2,732	2,629	2,521	2,407	2,286
28	9,283	6,440	5,317	4,607	4,299	4,019	3,811	3,648	3,518	3,411	3,246	3,072	2,889	2,794	2,694	2,591	2,483	2,369	2,246
29	9,229	6,395	5,276	4,659	4,252	3,983	3,774	3,612	3,483	3,376	3,211	3,037	2,855	2,759	2,660	2,556	2,447	2,333	2,210
30	9,179	6,354	5,238	4,623	4,227	3,949	3,741	3,580	3,450	3,344	3,178	3,005	2,823	2,727	2,627	2,594	2,415	2,299	2,176
40	8,827	6,066	4,975	4,373	3,986	3,712	3,508	3,349	3,222	3,116	2,953	2,781	2,589	2,502	2,401	2,295	2,183	2,063	1,931
60	8,484	5,795	4,729	4,139	3,760	3,491	3,291	3,134	3,008	2,904	2,741	2,570	2,387	2,289	2,187	2,078	1,962	1,834	1,688
120	8,179	5,539	4,497	3,920	3,548	3,248	3,087	2,933	2,808	2,705	2,543	2,372	2,188	2,089	1,983	1,870	1,746	1,605	1,431
∞	7,879	5,298	4,279	3,715	3,349	3,091	2,886	2,744	2,621	2,518	2,358	2,186	1,999	1,898	1,789	1,669	1,532	1,363	1,000

Оглавление

Курзенев В. А.

ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ДЛЯ УПРАВЛЕНЦЕВ

Учебное пособие
с примерами и задачами

Зав. редакционно-издательским отделом
Л. Е. Востряков
Редактор С. В. Чубинская-Надеждина
Корректор Е. В. Верецун
Верстка Е. И. Островой

Издательство СЗАГС. 199004, Санкт-Петербург, В. О. 8-я линия, д. 61.
Подписано в печать 31.01.2005 г. Формат 60×90/16. Усл.-печ. листов 13.
Печать офсетная. Гарнитура PragmaticaC, PetersburgC. Тираж 1000 экз.

Комплекс издательских и полиграфических работ выполнен ООО «Полиграфуслуги»
Издательский дом «Виктория плюс»