



Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Магнитогорский государственный технический университет им. Г. И. Носова»

Ю.В. Федосеева

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

*Утверждено Редакционно-издательским советом университета
в качестве учебного пособия*

Магнитогорск
2014

Рецензенты:

Кандидат физико-математических наук, доцент кафедры прикладной математики
и вычислительной техники,

ФГБОУ ВПО «Магнитогорский государственный университет»

О.А. Торшина

Преподаватель высшей категории Многопрофильного колледжа,
ФГБОУ ВПО «Магнитогорский государственный технический университет им. Г.И. Носова»

Е.А. Юрова

Федосеева Ю.В.

Теория вероятностей и математическая статистика [Электронный ресурс] : учебное пособие / Юлия Валерьевна Федосеева; ФГБОУ ВПО «Магнитогорский государственный технический университет им. Г.И. Носова». – Электрон. текстовые дан. (1,13 Мб). – Магнитогорск : ФГБОУ ВПО «МГТУ», 2014. – 1 электрон. опт. диск (CD-R). – Систем. требования : IBM PC, любой, более 1 GHz ; 512 Мб RAM ; 10 Мб HDD ; MS Windows XP и выше ; Adobe Reader 8.0 и выше ; CD/DVD-ROM дисковод ; мышь. – Загл. с титул. экрана.

Учебное пособие составлено в соответствии с рабочей программой по учебной дисциплине «Теория вероятности и математическая статистика» по специальности 230105, «Программное обеспечение вычислительной техники и автоматизированных систем» предназначено для студентов, обучающихся по очной форме обучения. Учебное пособие включает в себя теоретическую часть дисциплины.

УДК 519.2

© Федосеева Ю.В., 2014
© ФГБОУ ВПО «Магнитогорский
государственный технический
университет им. Г.И. Носова», 2014

Содержание

Введение	4
Раздел №1 Основы комбинаторики и теории вероятности.	6
Тема №1 Комбинаторика.	6
Тема №2 Основные понятия теории вероятностей.	12
Тема №3 Условная вероятность. Формула Байеса. Схема Бернули	22
Раздел №2 Случайные величины.	28
Тема №4 Случайная величина.	28
Тема №5 Виды распределений дискретной и непрерывной случайной величины.	44
Раздел №3 Основные понятия математической статистики.	55
Тема №6 Выборочное наблюдение. Ошибки выборки.	55
Тема №7 Статистические оценки параметров распределения.	66
Тема №8 Статистические гипотезы	71
Раздел №4 Метод статистических испытаний.	78
Тема №9 Моделирование случайных величин.	78
9.4 Моделирование случайной величины с биномиальным распределением.....	79
9.5 Моделирование случайной величины с равномерным распределением на отрезке $[a,b]$..	80
Тема №10. Статистический метод управления качеством.	85
Литература	93

ВВЕДЕНИЕ

Миссия данной дисциплины:

- формирование у будущего программиста основ знаний в области различных разделов математической статистики: комбинаторики, теории вероятностей, теории статистических испытаний;
- сформировать умения и навыки, позволяющие использовать компьютер в процессе решения задач с моделированием тех или иных случайных явлений.
- привить навыки сознательного и рационального использования компьютеров в своей учебной, а затем и профессиональной деятельности.

Знания, полученные при изучении дисциплины «Теория вероятности и математическая статистика» необходимые для изучения дисциплин: «Математические методы», «Численные методы», «Основы построения автоматизированных информационных систем».

Завершив обучение по курсу студенты **должны:**

иметь представление:

- о вероятностных закономерностях массовых случайных величин ;
- о способах сбора и группировки статистических сведений;
- о методах анализа статистических данных в зависимости от целей исследования;

знать:

- основные положения теории вероятности и математической статистики

уметь:

- использовать основные положения теории вероятности и математической статистики в профессиональной деятельности.

Входной тест

1. Закончите предложение. Значение производной функции... $f(x) = \frac{(2-4x)^2}{2}$ в точке $x_0=1$, равно...

- a) -18
- b) 16
- c) 8

2. Постройте график функции
$$f(x) = \begin{cases} 1 & \text{если } x < 0 \\ x & \text{если } 0 \leq x \leq 5 \\ -4 & \text{если } x > 5 \end{cases}$$

3. Вычислите без калькулятора:

$$\frac{e^0 + \sqrt{8}}{\cos(0) + 2\sqrt{4}}$$

4. Ответьте на вопрос: Петя, Вася и Таня садятся на 3 расположенных подряд стульев. Сколькими способами это можно сделать?

- a) 1
- б) 12
- в) 3
- г) 16

5. Вычислите: $4!$

- a) 4
- б) 24
- в) 8

6. Установите соответствие между парами множеств и соотношениями:

1. $A = \{a, b, c, d\}, B = \{a, c, d\}$ а). $A=B$

2. $A = \{\{a\}, a, b\}, B = \{a\}$ б). $A \subset B$

3. $A = \emptyset, B = \emptyset$ в). $A \supset B$

4. $A = \{a, d\}, B = \{d\}$ г). $A \in B$

5. $A = \{\{a, b\}, \{c, d\}\},$
 $B = \{\{a, b\}, \{c, d\}, c, d\}$ д). $B \in A$

7. Даны множества $A = \{0, 1, 2, 3, 4, 5, 6, 7\}, B = \{3, 4, 5, 6, 7, 8, 9\}, C = \{-3, -2, -1, 0, 1, 2, 3, 4\},$ Перечислите элементы множества $(A \cap B) \cup (C \cap D).$

8. Закончите предложение: Интеграл линейной функции $ax+b=f(x)$ имеет вид...

а) $\int (ax + b)dx = \frac{ax^2}{2} + bx + c$

б) $\int (ax + b)dx = a + 1 + c$

в) $\int (ax + b)dx = ax^2 + bx + c$

г) $\int (ax + b)dx = ax^2 + b + c$

9. Производная есть...

а) отношение двух функций;

б) отношение двух дифференциалов $\frac{df}{dx}$;

в) отношение дифференциала функции к независимой переменной $\frac{df}{x}$

10. Ответьте на вопросы: Заданы 4 переменные x, y, z, k . Каждая переменная может принимать только два значения: **истина, ложь**.

Сколько всевозможных комбинаций булевых значений имеют эти переменные?

а) 8

б) 16

в) 24

11. Какая из точек $A(2;-5), B(3;2), C(-4;1), D(-1;-2)$ расположена дальше всех от оси OX ...

а) В, С.

б) А, Д. $P=3$

в) А

12. Закончите предложение: Для суммирования диапазона значений вы используете следующие формулы:

а). $=СУММ(C22+C23+C25+C26)$

б). $=СУММ(C22..C26)$

в). $=(C22..C26)$

г). $=СУММ(C22:C26)$

13. Ответьте на вопрос Какая из формул определяет средние значения?

а). $=(C22...C26)/5$

б). $=(C22+C23+C24+C25+C26)/5$

в). =ССРЕДЗН(C22:C26)/5

г). =СРЕДЗН(C22:26)

14. Укажите верную запись ссылки на ячейку В18 листа 3 текущей книги, при работе с листом 5.

а).Лист 3, В18.

б).Лист 3! В18.

в).Лист 3; В18.

15. Ответьте на вопрос: Как будет вычисляться формула, содержащая синтаксическую ошибку?

а).В ячейке будет выведено сообщение об ошибке;

б).Результатом вычисления будет 0;

в).Система не допускает ввода формул с синтаксическими ошибками;

г).В ячейке будет выведено #####. P=4

РАЗДЕЛ №1 ОСНОВЫ КОМБИНАТОРИКИ И ТЕОРИИ ВЕРОЯТНОСТИ

Цели изучения: сформировать умения и навыки описания, объяснения, выявление закономерностей явлений действительности на основе рассматриваемых понятий и законов.

После изучения раздела Вы будете:

Иметь представление:

- о комбинаторики;
- о теории вероятности как науке;
- о вероятностных закономерностях массовых случайных величин.

Знать:

- основные понятия комбинаторики (размещения, перестановки, сочетания);
- основные понятия теории вероятностей;
- формулы полной вероятности, Байеса, Бернулли;

Уметь:

- определять вид события;
- использовать основные понятия и формулы теории вероятности при решении задач;

Тема№1 Комбинаторика.

Изучив эту тему, Вы будете:

Иметь представление:

- о комбинаторики;

Знать:

- основные понятия комбинаторики (размещения, перестановки, сочетания);

Уметь:

- использовать основные понятия и формулы при решении комбинаторных задач;

Введение

Случай, случайность – с ними мы встречаемся повседневно: случайная встреча, случайная находка, случайная ошибка. Казалось бы, тут нет места для математики, но и здесь наука обнаружила закономерности. О статистике слышали все, но вместе с тем едва ли найдется еще одна наука, в которой так горячо спорили бы о ее предмете, то есть о том, что она собой представляет.

Итак, что же представляет собой термин «статистика»? Это слово многозначно. В настоящее время насчитывается около тысячи определений статистики. Первое из них относиться к

1749 году: на протяжении 250 лет оно уточнялось и дополнялось. Сначала статистику определяли как

«Staten Kunde» - государствоведение (описание достопримечательностей государств).

В настоящее время под термином «статистика» чаще всего понимают следующее.

Статистика- одна из общественных наук, имеющая целью сбор, упорядочивание, анализ и сопоставление числового представления фактов, относящихся к самым разнообразным массовым явлениям.

Статистика- это также одна из форм практической деятельности людей, цель которой- сбор, обработка и анализ массовых данных о тех или иных явлениях.

Статистикой называют также **различного рода числовые или, как часто говорят, цифровые данные**, характеризующие различные стороны жизни государства: политические отношения, культуру, население, производство и так далее.

Часто слово статистика употребляется в качестве более короткого эквивалента для слов «**статистические методы**». Статистические методы можно охарактеризовать как методы, применяемые при сборе, представлении, анализе и интерпретации данных. В качестве примера можно упомянуть о методах, применяемых при сборе данных о совокупности студентов колледжей, обработке этих данных, обобщении и представлении в виде различных итоговых абсолютных, относительных и средних показателей с помощью графиков, таблиц.

Если обратиться к истории статистики, то ее развитие сходно с развитием языка и счета. Эта наука имеет древние корни. Вот лишь некоторые сведения: в Китае более чем за две тысячи лет до нашей эры производились исчисления населения по полу и возрасту, а также собирались сведения о состоянии промышленности и сельского хозяйства. Упоминание о статистических обследованиях встречаются и в библейские времена. В Древнем Риме велась статистика численности населения и имущественного положения граждан.

Однако если собирание статистических данных началось в самой глубокой древности, то их обработка и анализ, то есть зарождение **статистики- науки**, относятся к более позднему периоду- второй половине XVII века.

Гораздо ближе к современному пониманию статистики была английская школа политических арифметиков. Основателями школы были Д. Граунт (1620-1674), Э. Галлей (1656-1742), В. Пети (1623-1687). В этих трудах наметилось два направления: **демографическое** с преобладанием вопросов страхования жизни и **статистико-экономическое**.

В первой половине XIX века возникло третье направление статистической науки – **статистико-математическое**. Особый вклад в развитие этого направления внес бельгийский статистик А. Кетле (1796-1874). Двумя другими учеными, внесшими вклад в развитие статистики, являются два англичанина – В. Гальтон и К. Пирсон. Представители статистико-математического направления считают основой статистики теорию вероятностей, составляющую одну из отраслей прикладной статистики. Поэтому два из четырех модульных блоков посвящены теории вероятностей. Но вернемся к понятию математическая статистика.

Математическая статистика- наука, изучающая методы обработки результатов наблюдений случайных массовых явлений, обладающих статистической устойчивостью, закономерностью, с целью выявления этой закономерности.

Итак, статистика изучает числа, чтобы обнаружить в них закономерности. Закономерности составляют главный объект научных исследований. Например, исследователи могут интересоваться вопросами типа: как изменится давление в жидкости с изменением глубины? С какой скоростью движутся падающие тела и т.п. Знание законов природы позволяют ответить на подобные вопросы, не производя реальных опытов, т.е. заранее. Например, можно точно вычислить, какие вещества и в какой пропорции образуются при той или иной химической реакции, или предсказать, когда в данной местности произойдет следующее солнечное затмение.

Но не во всех ситуациях интересующий нас результат полностью и жестко определяется влияющими на него факторами. Например, нельзя указать, сколько часов будет светить электрическая лампочка. Невозможно предвидеть число посетителей магазина и количество то-

варов, которое они купят. Ответ на подобные вопросы можно получить, только проведя соответствующие испытания. Часто явления (ситуации), в которых результат полностью определяется влияющими на него факторами, называются **детерминированными** или **закономерными**, а те, в которых это не выполняется – **недетерминированными** или **стохастическими**. В большинстве явлений присутствуют оба вида изменчивости – и закономерная, и случайная, и для нахождения закономерностей нам приходится «отсеивать» мешающие случайные факторы. Изучением закономерностей, которые порождаются случайными событиями, занимается наука теория вероятностей. Изучение теории вероятности невозможно без знаний математики, в частности одного из разделов математики – комбинаторика.

1.1 Комбинаторика.

При решении многих практических задач приходится выбирать из некоторой совокупности объектов элементы, обладающие тем или иным свойством, располагать эти элементы в определенном порядке. Поскольку в таких задачах речь идет о тех или иных комбинациях объектов, их называют комбинаторными задачами. Область математики, в которой изучаются комбинаторные задачи, называют **комбинаторикой**.

Определение: Комбинаторика – раздел математики изучающий количество комбинаций, подчиненных определенным условиям, составленных из элементов конечного множества.

Пусть имеется некоторое множество, содержащее конечное число членов. Например, множество учебных групп в колледже, множество книг на полке, множество населенных пунктов в данной области или, например, множество целых положительных чисел, меньших 10, и т.д. Все элементы этого множества можно пронумеровать, в результате получается некоторая последовательность элементов данного множества, которые обычно записываются в виде $a_1, a_2, \dots, a_n$. Такие «занумерованные» множества будем называть **комбинациями**. Если в этом множестве поменять местами хотя бы два элемента, то получим новую комбинацию, которой будет соответствовать новая последовательность элементов данного множества.

Определение: Перестановками называют комбинации, состоящие из одних и тех же n различных элементов и отличающиеся только порядком их расположения.

$$P_n = n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 1. \quad (1)$$

Пример №1: Сколько трехзначных чисел можно составить из цифр 1, 2, 3, если каждая цифра входит в изображение числа только один раз.

Решение.

Подсчитаем вручную: (123) (132) (213) (231) (312) (321)

Или с помощью формулы (1): $P = 3! = 3 \cdot 2 \cdot 1 = 6$;

Если в последовательности $a_1, a_2, \dots, a_n$ не только менять местами элементы, но и заменять последовательно элементы на другие $b_1, b_2, \dots, b_n$, то полученные новые комбинации можно подсчитать с помощью формулы размещения.

Определение: Размещения называют комбинации, составленные из n различных элементов по m элементов, которые отличаются либо составом элементов, либо их порядком.

$$A_n^m = n(n-1)(n-2)\dots(n-m+1) = \frac{n!}{(n-m)!} \quad (2)$$

Пример №2: Сколькими способами можно выбрать из группы, насчитывающего 40 студентов, старосту, зам. старосты, физорга.

Решение.

Любой такой выбор является размещением без повторений из 40 элементов по 3. Значит, число способов выбора равно $A_{40}^3 = 40 \cdot (40-1) \cdot (40-2) = 40 \cdot 39 \cdot 38 = 59280$.

Пример №3: Сколько можно составить сигналов из 6 флажков различного цвета, взятых по 2?

Решение.

Используя формулу (2), получим: $A_6^2 = 6 \cdot (6 - 1) = 6 \cdot 5 = 30$ сигналов.

Если при замене элемента a_i на b_i порядок расположения b_i не важен, то количество полученных комбинаций подсчитывается по формуле сочетания.

Определение: Сочетания называют комбинации, составленные из n -различных элементов по m элементов, которые отличаются хотя бы одним элементом.

$$C_n^m = \frac{n!}{m!(n-m)!}; \quad (3)$$

Необходимо отметить, что: $C_n^0 = C_n^{n-n} = 1 = C_n^n$; $C_n^m = C_n^{n-m}$.

Пример №4: В аудитории имеется 10 лампочек. Сколько существует разных способов ее освещения, при которых горит ровно 2 лампочки?

Решение.

Способов освещения столько, сколько существует сочетаний из 10 лампочек по 2, то есть

$$C_{10}^2 = \frac{10!}{(2!8!)} = \frac{10 \cdot 9 \cdot 8!}{2! \cdot 8!} = \frac{90}{2 \cdot 1} = 45;$$

Пример №5: Из множества $\{a, b, c, d, e\}$ можно составить 10 сочетаний по 3 элемента в каждом:

$\{a, b, c\}, \{a, b, d\}, \{a, b, e\}, \{a, c, d\}, \{a, c, e\}, \{a, d, e\},$
 $\{b, c, d\}, \{b, c, e\}, \{b, d, e\}, \{c, d, e\}$

Из каждого такого сочетания путем различного упорядочивания элементов получается 6 размещений из 5 элементов по 3. Например, из сочетания $\{a, b, c\}$ получаем следующие размещения $(a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a)$. Отсюда видно, что число размещений без повторений из 5 элементов по 3 равно $6 \cdot 10 = 60$, что согласуется с формулой $A_5^3 = 5 \cdot 4 \cdot 3 = 60$.

Различные сочетания порождают различные размещения и каждое размещение может быть получено указанным способом. Иными словами $A_n^m = P_m \cdot C_n^m$.

Задание 1.

Запишите для указанных задач в таблицы формулы для расчета количества искомых комбинаций.

Текст задачи	Формула для вычисления
В группе из 15 студентов нужно выбрать троих для поездки в Германию. Сколькими способами это можно сделать?	
В группе из 15 студентов нужно выбрать троих для продолжения учебы во Франции, Англии и США. Сколькими способами это можно сделать?	
В группе из 10 юношей и 12 девушек необходимо выбрать 5 юношей и 5 девушек для проведения анкетирования в колледже. Сколькими способами это можно сделать?	

До сих пор мы рассматривали соединения, в каждом из которых любой из n различных элементов мог входить не более одного раза. Но в некоторых случаях элементы в соединении могут повторяться. Например, любое трехзначное число можно рассматривать как соединение из трех цифр, причем цифры могут повторяться: 771, 722, 545, 333 и т.п.

Пусть имеется m групп элементов (в каждой группе достаточно много элементов), таких, что элементы внутри группы неразличимы между собой, а элементы разных групп различимы. Из совокупности всех элементов возьмем подмножество, содержащее n элементов. Это подмножество из n элементов определяется числом взятых элементов из 1-ой группы, числом взятых элементов из 2-ой группы, и т.д. Эти числа могут принимать значения от 0 до n , но так, чтобы сумма их равнялась n . Число различных способов образования n -элементного множества находится по так называемой формуле числа различных сочетаний с повторениями

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

Пример №6. Имеются пирожные 7 различных типов. Пирожные одного и того же типа считаем неразличимыми. Сколько существует различных способов покупки 12 пирожных?

Решение. В нашем случае число групп $m = 7$. Пирожные, относящиеся к разным типам, различимы, пирожные, относящиеся к одному типу, неразличимы. Рассмотрим множество из $n = 12$ покупаемых пирожных. Способ образования такого множества определяется числом купленных пирожных 1-го типа, 2-го типа, ..., 7-го типа. Числа эти неотрицательные, целые и в сумме дают 12, т.е. мы имеем дело с сочетаниями с повторениями, где $m = 7$, $n = 12$. Поэтому число различных способов покупки 12 пирожных находим по формуле:

$$C_{n+m-1}^{m-1} = C_{12+7-1}^{7-1} = C_{18}^6 = \frac{18!}{6!12!}.$$

Пусть имеются k групп элементов, причем в первой группе n_1 неразличимых элементов, во второй группе n_2 неразличимых элементов, ..., в k -ой группе – n_k неразличимых элементов. Элементы из разных групп различимы. Таким образом, имеем всего $n_1 + n_2 + \dots + n_k = n$ элементов. Рассмотрим всевозможные упорядоченные наборы из этих элементов. Число различных таких наборов (*перестановок с повторениями*) равно

$$\frac{(n_1 + n_2 + \dots + n_k)!}{n_1!n_2!\dots n_k!}.$$

Пример №7. На 6-ти карточках написаны буквы, из которых можно составить слово АНАНАС. Сколько существует различных шестибуквенных слов, которые можно составить при помощи этих 6-ти карточек?

Решение. 6 карточек разобьем на 3 группы. Первая группа образована карточками с буквой А. Число таких карточек равно 3. Они неразличимы по буквам, у всех одна и та же буква А, $n_1 = 3$. Вторая группа образована двумя карточками, содержащими букву Н. Элементы второй группы также неразличимы между собой, $n_2 = 2$. Третья группа образована одной карточкой с буквой С, $n_3 = 1$. Таким образом, мы имеем дело с перестановками с повторениями и число слов из 6-ти букв равно

$$\frac{(3+2+1)!}{3!2!1!} = \frac{6!}{3!2!} = 60.$$

Задание 2.

Запишите для указанных задач в таблицы формулы для расчета количества искомых комбинаций.

Текст задачи	Формула для вычисления
На 10-ти карточках написаны буквы так, что из этих карточек можно составить слово МАТЕМАТИКА. Сколько существует различных 10-буквенных слов, которые можно образовать при помощи этих десяти	

<i>карточек?</i>	
<i>На почте имеются марки 10-ти различных типов. Покупается 15 марок. Сколько существует различных способов покупки 15-ти марок?</i>	
<i>На полке наудачу располагаются 10 книг. Сколько существует различных способов расположения 10-ти книг?</i>	
<i>На полке наудачу располагаются 10 книг. Сколько существует различных способов расположения 10-ти книг, при которых 2 заранее помеченные книги окажутся рядом?</i>	

Задания:

1. Имеется множество цифр {3,5,7}. Сколькими способами можно расставить цифры, чтобы получить различные числа? Запиши полученные числа
2. Необходимо выбрать в подарок 4 из 10 имеющихся книг. Сколькими способами можно это сделать?
3. В некотором журнале 10 страниц. Необходимо на страницах этого журнала поместить 5 фотографий. Сколькими способами это можно сделать, если ни одна страница не должна содержать более одной фотографии?
4. Сколькими способами можно составить команду из 4 человек для соревнований по бегу, если имеются 7 бегунов?
5. Десять человек решили обменяться фотографиями. Сколько нужно фотографий?
6. Сколько пятизначных номеров можно составить из девяти цифр 1,2,3,4,5,6,7,8,9?
7. Из вершины прямого угла проведены внутри 5 лучей. Сколько острых углов при этом образовались?
8. Сколько пятизначных номеров можно составить из цифр 2,3,5,7?
9. Сколькими способами можно поставить 8 шашек на черные поля доски?
10. У одного человека есть 11 книг по математике. Сколько пятизначных номеров можно составить из девяти цифр 1,2,3,4,5,6,7,8,9?, у другого – 15 книг. Сколькими способами они могут выбрать по 3 книги каждый для обмена?
11. Имеется 10 белых и 5 черных шаров. Сколькими способами можно выбрать 7 шаров, чтобы среди них были 3 черных?
12. На прямой 5 точек: А,В,С,Д,Е. Сколько получится отрезков?
13. В партии из 10 дискет 7 стандартных. Найти вероятность того, что среди 6 взятых наудачу дискет 4 стандартные.
14. Имеется 20 мониторов из них 2 неисправных остальные в рабочем состоянии. Сколькими способами можно выбрать 7 мониторов, чтобы среди них были 2 неисправных?

Тест для самоконтроля

1. Ответьте на вопросы. Сколькими способами 6 человек могут сесть на 6 стульев?
 - а). 120
 - б). 720
 - в). 36
2. Запиши формулу для решения следующей задачи: Задано число 12345. Сколько чисел начинается с 12?
 - а). $A_5^3 \cdot A_5^2$
 - б). P_5
 - в). P_3

2.1 События

Теория вероятности возникла как наука из убеждения, что в основе массовых случайных событий лежат детерминированные закономерности. Теория вероятности изучает данные закономерности.

Например: определить однозначно результат выпадения “орла” или “решки” в результате подбрасывания монеты нельзя, но при многократном подбрасывании выпадает примерно одинаковое число “орлов” и “решек”.

Издавна люди пытались выяснить закономерности выпадения выигрыша в азартных играх. Есть игры, в которых от игроков не требуется никакого умения. Например, «орлянка»: монету подбрасывают, и в зависимости от того, какой стороной она упала, выигрывает первый или второй партнер. Или другая игра, где властвует случай, - игра в кости. Когда игральную кость бросают, то на ее верхней грани оказывается какое-то число – от 1 до 6, по числу граней. Предположим, что кубик совершенно правильный – все его грани абсолютно одинаковы. Если подбрасывать кубик 10 раз, то числа могут выпасть, например, в такой последовательности: 1, 5, 2, 2, 3, 4, 1, 6, 3, 5 или в такой: 6, 5, 5, 6, 3, 1, 1, 4, 2, 3, 6. В первом случае 6 выпала один раз, а во втором – три раза. Может случиться, что шестерка не выпадет ни разу или все 10 раз.

А если подбрасывать кубик не 10, а 10.000 раз, тогда каждая грань будет выпадать примерно одинаково часто. Пусть N_k – количество выпадений грани с цифрой k при N бросаниях кубика, тогда N_k / N называется частотой выпадения грани k .

Поскольку $N_1 + N_2 + N_3 + N_4 + N_5 + N_6 = N$, то все частоты будут приблизительно одинаковы и равны $1/6$. Сама величина $1/6$ называется **вероятностью** выпадения как 5 или 6 или 4 и т.д. Слово «**событие**» в быту применяют к значительным явлениям, а в математике – ко всем возможным исходам рассматриваемой ситуации.

Определение: Под **событием** будем понимать все то, что может произойти, а может и не произойти. Событие – не происшествие, а лишь возможный исход опыта или наблюдения.

События обозначают заглавными латинскими буквами: $A, B, C, \dots$

Например: Стрелок стреляет по мишени, разделенной на 4 области.

Выстрел – *испытание*

Стрелок стрелял, и попал – *событие*

Как уже отмечалось ранее наблюдаемые нами события можно подразделить на 2 вида:

- детерминированные явления;
- статистические (случайные) явления.

К детерминированным (предопределенным) явлениям относятся достоверные и невозможные события.

К статистическим явлениям относятся случайные события.

Достоверными называют события, которые обязательно произойдут, если будет осуществлена определенная группа условий.

Невозможными называют события, которые заведомо не произойдут, если будет осуществлена совокупность условий.

Случайными называют события, которое при осуществлении совокупности условий, могут произойти или не произойти.

Пример №1:

- выпадение решки при бросании монеты является случайным событием, так как условия или причины появления этого события различны: форма монеты, сила с которой брошена монета и т.д;
- при подбрасывании кубика невозможное событие - кубик станет на ребро, случайное событие - выпадение какой либо грани.

События могут быть **равновозможными**. Равновозможность означает равноправность определенных исходов испытания относительно некоторого комплекса условий.

Пример №2: В ящике находятся белая и серая мыши. Считаем, что они одинаково голодные и подвижные. На запах сыра из ящика выбегает одна мышь. Это событие равновозможное относительно двух исходов. А так же несовместные.

Определение: Два события называются **несовместными**, если наступление одного из них исключает возможность наступления другого. В противном случае события называются совместными.

Пример №3: Несовместные события: выпадения орла и решки при бросании монеты; Событие, состоящие в том, что в результате бросания кости выпадет 1,2,3,4,5 и 6 очков.
Совместные события: выпадение числа "2" и выпадение четного числа при бросании игральной кости.

Определение: Несколько событий образуют **полную группу**, если в результате испытания появиться хотя бы одно из них.

Пример №4: Приобретены два билета денежно-вещевой лотереи. Обязательно произойдет одно из следующих событий:

- | | | |
|--|---|---------------------------|
| а) выиграет 1-й билет
б) выиграет 2-ой билет
в) выиграют оба
г) не выиграет ни один | } | полная группа
событий. |
|--|---|---------------------------|

В результате испытания происходят только одно из элементарных событий. Совокупность всех возможных, различных исходов испытаний называется **пространством элементарных событий**.

Пример №5: Испытание - подбрасывание шестигранного кубика. Элементарное событие - выпадение грани с "1" или "2".

Задание 1. Выделите события и укажите их виды:

- Восемь различных книг расставлены наудачу на одной полке. Найти вероятность того, что две определенные книги окажутся поставленными рядом
- В ящике 10 деталей, среди которых 2 бракованных. Найти вероятность того, что среди наудачу отобранных 6 деталей окажется не более одной бракованной детали.
- Из семи лотерейных билетов два выигрышных. Семь человек по очереди берут по билету (не возвращая его обратно). Чему равна вероятность выигрыша и зависит ли она от места в очереди?
- Игральная кость подбрасывается два раза. Какова вероятность того, что сумма выпавших очков = 4?

Теория вероятностей не ставит перед собой задачу предсказать, произойдет событие единичное или нет. Но при многократном появлении события, появляются определенные закономерности. Так вот установлением этих закономерностей занимается теория вероятностей.

Теория вероятностей – это раздел математики, который возник в XVII веке как способ выяснения при каких условиях можно получить максимальный выигрыш в азартных играх (карты, кости, рулетка).

Предметом теории вероятностей является изучение вероятностных закономерностей массовых однородных случайных событий.

Впервые вероятности случайных событий в азартных играх вычислили французские математики Блез Паскаль и Пьер Ферма. Они подсчитали число шансов события из общего возможного числа равновероятных исходов.

Можно выделить несколько определений вероятностей.

1. Интуитивная вероятность.
2. Классическое определение вероятности (частный случай).

3. Статистическое и геометрическое определение вероятности.
4. Аксиоматическое определение вероятности.

2.2 Классическая вероятность

Определение: Классической вероятностью называется отношение числа благоприятных исходов для заданного события к общему количеству всех возможных исходов данного опыта.

$$P(a) = \frac{m}{n}, \quad n - \text{общее количество исходов}$$

m – число благоприятных исходов.

Из определения вероятности вытекают следующие свойства:

Свойство 1. Вероятность достоверного события = 1.

Свойство 2. Вероятность невозможного события = 0.

Свойство 3. Вероятность случайного события есть положительное число, заключительное между 0 и 1 то есть $0 \leq P(A) \leq 1$.

В простых случаях вычислить вероятности не составляет труда.

Например, если требуется найти вероятность выпадения "орла" при однократном бросании монеты, то решение будет выглядеть следующим образом:

$A = [\text{событие появления "орла"}]; m = 1; \quad n = 2;$

$$P(A) = \frac{1}{2} = 0,5;$$

Задание 2: Определите и запишите вероятность появления карты червовой масти, туза из колоды в 36 карт.

События $A = [\text{вынута карта червонной масти}], B = [\text{вынут туз}]$

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

В некоторых случаях вычислить вероятности событий отнюдь не просто.

Пример 4

Пусть событие B_k означает, что при бросании 2 игральных костей сумма выпавших очков равна k . Всего имеется 11 подобных событий: $B_2, B_3, \dots, B_{12}$.

Однако вероятности их неодинаковы. Равновероятным исходам соответствует пары чисел (i, j) , где i – число выпавших очков на 1-й кости, j – число выпавших очков на 2-ой кости. Всего таких исходов 36.

Шансами события B_k будут исходы при которых $i+j=k$.

Число шансов каждого из них можно подсчитать с помощью таблицы.

В таблице каждая строка соответствует числу очков i на 1-ой кости, каждый столбец – числу очков j на 2-й кости, а в клетке на пересечении i и j стоит сумма $i+j=k$. Все исходы (i, j) – равновероятны – их 36, а число шансов события B_k равно количеству клеток таблицы с числом k . Поэтому, по классическому определению вероятности, получаем:

$i \backslash j$	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$$P(B_2) = P(B_{12}) = \frac{1}{36}; \quad P(B_3) = P(B_{11}) = \frac{2}{36} = \frac{1}{18}; \quad P(B_4) = P(B_{10}) = \frac{3}{36} = \frac{1}{12};$$

$$P(B_5) = P(B_9) = \frac{4}{36} = \frac{1}{9}; \quad P(B_6) = P(B_8) = \frac{5}{36}; \quad P(B_7) = P(B_7) = \frac{6}{36} = \frac{1}{6}.$$

Рассмотренный пример является частным случаем, так называемой «урновой» схемы. Пусть в ящике (урне) лежит g шаров. Они различаются по цвету, или метками, но не на ощупь. Множество шаров $U = \{u_1, u_2, \dots, u_n\}$. Пусть A любое подмножество $A \subset U$.

Тогда вероятность события A равна отношению $P(A) = \frac{|A|}{|U|}$, где $|U|$ - общее число шаров в урне; $|A|$ - число шаров в подмножестве A .

2.3 Операции над событиями

Определение: Объединением, или суммой $A + B$ двух событий A и B называют событие, состоящее в появлении события A или события B , или обоих этих событий.(рис.1)

Обозначение: $C = A \cup B$ или

$$C = A + B$$

Пример №5: Если из оружия произведены два выстрела, события: A – попадание при первом выстреле; B – попадание при втором выстреле, тогда $A+B$ – попадание при первом, или

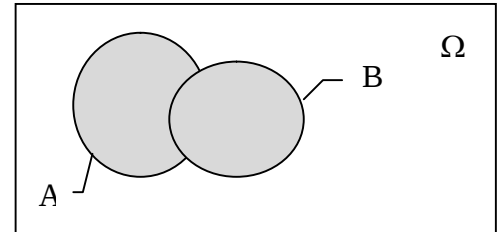


рис.1

при втором или в обоих выстрелах.

Теорема: Вероятность появления одно из двух несовместимых событий, безразлично какого, равна сумме вероятностей этих событий: $P(A + B) = P(A) + P(B)$.

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

Следствие из теоремы:

Вероятность появления одного из нескольких попарно несовместимых событий, безразлично какого, равна сумме вероятностей этих событий:

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

Задание 3: Найдите и запишите вероятность появления цветного шара если в урне 30 шаров: 10 красных, 5 синих и 15 белых.

$A =$ _____,

$B =$ _____

$P(A) = \frac{?}{?}$ $P(B) = \frac{?}{?}$ $P(A + B) =$ _____

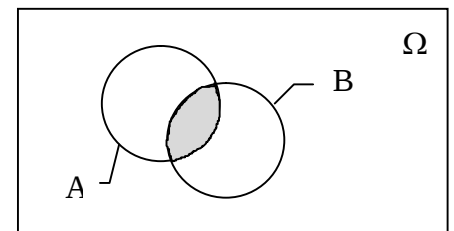


Рис.2

Определение: Пересечением или произведение AB событий A и B называют событие, которое происходит тогда и только тогда, когда имеют место оба события A и B . (рис.2)

Обозначение: $C = A \cap B$ или $C = A * B$

Пример №6: Событие $C = [\text{вынут червонный туз}]$ является произведением событий A и B , где A – вынута карта червонной масти, а B – вынут туз.

Определение: Отрицанием события A называют такое событие, которое состоит в том, что A не происходит. Обозначение для него $\bar{A}$.

Определение: Противоположными называют два единственно возможных события образующие полную группу. (рис.3)

Пример №7: При стрельбе по мишени событие $A = [\text{попадание в мишень}]$, а событие $\bar{A} = [\text{промах}]$.

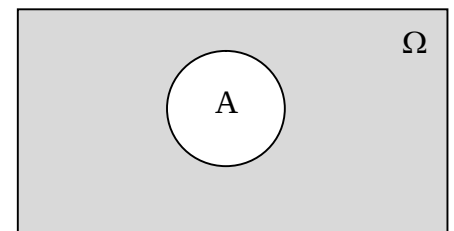


Рис.3

Теорема: Сумма вероятностей противоположных событий равна единице. $P(A) + P(\bar{A}) = 1$.

Доказательство: Противоположные события образуют полную группу, а сумма вероятностей событий, образующих полную группу, равна единице.

Пример №8: Консультационный пункт института получает пакеты контрольных работ из городов А, В, С. Вероятность получения пакета из города А=0,7 из В=0,2. Найти вероятность того, что очередной пакет будет получен из города С.

Решение: События А, В, С – образуют полную группу событий.

$$1 - P(A) - P(B) = P(C)$$

$$P(C) = 1 - 0,7 - 0,2 = 0,1.$$

Замечание: Если вероятность одного из двух противоположных событий обозначена через p , то вероятность другого события обозначают через q . Таким образом, в силу предыдущей теоремы $p+q=1$.

Недостатки классического определения вероятности:

1. Невозможность в некоторых случаях представить результат испытаний в виде совокупности событий (либо благоприятных, либо всех испытанных). Например, подсчитать, с какой вероятностью у женщины родится девочка или мальчик?
2. Иногда трудно указать основания позволяющие считать элементарные события равновероятными. Например, для определения вероятности появления события при бросании игральной кости – предполагается, что это куб изготовлен из однородного материала.

2.4 Статистическая и геометрическая вероятность

И еще один подход к понятию вероятности, не связанный с числом всех исходов и благоприятных исходов. В основе этого подхода лежит **явление устойчивости** частоты поступления события при многократном повторении испытаний. (т.е. кость бросают не один раз и находят классическую вероятность, а n -ое число раз и подсчитывают сколько раз выпала цифра 6 или 5).

Пусть в результате некоторого опыта возможно появление события А. Повторим опыт n раз и подсчитаем количество m наступлений событий А. Будем предполагать, что в данной серии опытов результаты предшествующих испытаний не на последующие влияют.

n – увеличиваясь мало изменяет отношение m/n , поэтому если монету подбросить многократно, то отношение количества выпавших гербов к числу всех исходов приближенным будет равно $1/2$.

Определение: Отношение m/n – называется частотой появления события А в n испытаниях.

$$W(A) = \frac{m}{n} \quad \text{где } 0 \leq \frac{m}{n} \leq 1$$

m – число испытаний в которых событие появилось;

n – общее число фактически произведенных испытаний.

Задание 4 Запишите в таблицу отличия и сходства с классическим определением вероятности:

Классическая

Статистическая

Пример №9: Отдел технического контроля обнаружил 3 нестандартные детали в партии из 80 случайно отобранных деталей. Найти относительную частоту появления нестандартных деталей.

Решение:
$$W(A) = \frac{3}{80};$$

Свойство статистической устойчивости вероятности: в различных испытаниях частота изменяется мало тем меньше, чем больше произведено испытаний, колеблясь около некоторого постоянного числа.

Пример №10: По данным шведской статистики частота рождения девочек за 1999 г по месяцам характеризуется следующими числами

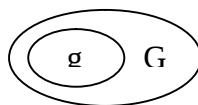
(0,486; 0,489; 0,490; 0,471; 0,478; 0,482; 0,462; 0,484; 0,485; 0,491; 0,482; 0,473). Тогда $W \approx 0,482$ – приближенное значение колеблется около вероятности рождения девочек;

Недостатки статистического определения вероятности: Неоднозначность статистической вероятности, так в нашем примере в качестве вероятности можно взять: 0,482; 0,491; 0,472; 0,481; и т.д.

Чтобы преодолеть недостаток классического и статистического определения вероятности, состоящее в том, что они неприменимы к испытанию с ∞ числом исходов вводят – **геометрическую вероятность попадания точки в область (отрезок, часть плоскости и т. д.)**

Пусть, например, стержень наудачу разламывается на три части. Какова вероятность того, что из получившихся отрезков можно будет построить треугольник?

В этой задаче мы имеем бесконечное множество исходов, так как разлом может попасть на любую точку стержня. Поэтому данные выше определения вероятности события не годятся. Будем использовать геометрическое определение вероятности. Разберем следующую модель. Пусть на отрезке СД бросают на удачу точку. Пусть отрезок СД составляет часть отрезка АД. На отрезке АД наудачу поставлена точка. Найти вероятность, что точка попадает на отрезок СД. Назовем вероятностью части этого отрезка отношение длины этой



часть этого отрезка отношение длины этой

часть к длине всего отрезка.

$$P = \frac{\text{Длина СД}}{\text{Длина АД}} ; \text{ если фигуры плоские, то}$$

$$P = \frac{\text{площадь } g}{\text{площадь } G}$$

Для геометрической вероятности справедливы следующие утверждения:

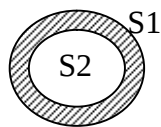
1. Для любой части фигуры значение вероятности является неотрицательным числом, не превосходящим 1. Для самой фигуры значение вероятности равно 1.
2. Если части X и Y некой фигуры не имеют общих точек (несовместны), то $P(X \cup Y) = P(X) + P(Y)$.

На основе этих утверждений для геометрической вероятности можно определить те же понятия, что и в случае конечного вероятностного пространства, доказать аналоги формул сложения и умножения.

Пример №11

На плоскости начерчены две концентрические окружности радиусы которых 5 и 10 см соответственно. Найти вероятность что точка, брошенная наудачу в большой круг, попадет в кольцо образованное окружностями.

Решение. $A = [\text{событие состоящее в том, что точка попадет в кольцо образованное окружностями}]$.

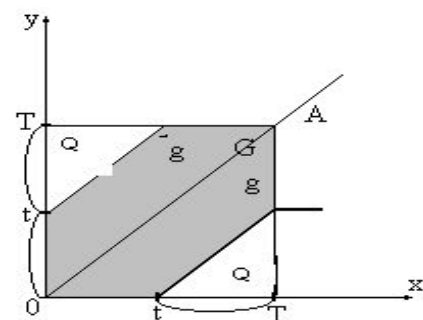


$$P = \frac{S1 - S2}{S1} = \frac{\pi r_1^2 - \pi r_2^2}{\pi r_1^2} = \frac{\pi(10^2 - 5^2)}{\pi \cdot 10^2} = \frac{100 - 25}{100} = \frac{75}{100};$$

Пример №12

В сигнализатор поступают сигналы от двух устройств, причем поступление, каждого из сигналов равновозможно в любой момент промежутка времени длительность от 0 до T. Моменты поступления сигналов независимы один от другого. Сигнализатор срабатывает, если разность между моментами поступления сигналов меньше t ($t < T$). Найти вероятность того, что сигнализатор срабатывает за время T, если каждое из устройств пошлет по одному сигналу.

Решение:



Обозначим моменты поступления сигналов через x и y, тогда в силу условия задачи:

$$0 \leq x \leq T \quad \text{и} \quad 0 \leq y \leq T$$

Введем прямоугольную систему координат. Обозначим этот квадрат за фигуру G.

Число всех исходов это подача всевозможных сигналов в любой момент времени от 0 до T. Благоприятные исходы: сигнализатор срабатывает, если разность между моментами поступления меньше t.

$x - y < t$ (в случаях $x > y$)

$y - x < t$ если $x < y$

Отметим это на графике. Обозначим прямую $x = y$, когда $t=0$,

$y < t + x$, когда $x < y$, $y > x - t$, когда $x > y$.

Тогда I область выше $x = y$ и ниже $y = t + x$.

II область ниже $x = y$ выше $y = x - t$.

Площадь области g можно определить по формуле $S_g = S_G - S_Q = T^2 - (T - t)^2$. Тогда вероятность того, что сигнализатор сработает за время T, равна:

$$P = \frac{S_g}{S_G} = \left(\frac{T^2 - (T - t)^2}{T^2} \right) = \frac{T^2 - T^2 + 2Tt + t^2}{T^2} = \frac{t^2 + 2tT}{T^2}.$$

2.5 Аксиоматическое определение вероятности

Аксиома – исходное утверждение предложения какой-либо нужной теории, которое принимается в качестве недоказанного в данной теории и из которой выводятся все остальные предложения.

Для введения аксиом нужны соглашения. Зафиксируем комплекс условий. Рассмотрим некоторую систему S, A, B, C и т.д., каждая из которых при осуществлении комплекса условий должно произойти или не произойти.

Введем соглашения:

1. Если при каждом осуществлении команды условий, при котором происходят события A, происходит событие B, то мы будем говорить: A влечет за собой B, и обозначают символом $B \supset A$ (из B следует A), $B \subset A$ (из A следует B).
2. Если A влечет за собой B и в тоже время B влечет A, т.е. если при каждой реализации комплекса условий события A и B наступают или не наступают одновременно, то события A и B равносильны, или A и B тождественны. $A = B$.
3. Событие состоящие из наступления обоих событий A и B будем называть произведением событий A и B ($A \cap B$).
4. Событие состоящие хотя бы из одного A и B будем называть суммой A + B ($A \cup B$).
5. События, состоящие в том, что событие A происходит, а событие B не происходит, называется разностью событий A и B $A - B$ ($A \setminus B$) – дополнением до B.
6. Два события A и $\bar{A}$ называются противоположными, если для них одновременно выполняются два соотношения:
 $A + \bar{A} = 1$ - достоверное событие; $A * \bar{A} = 0$ - невозможное событие.
7. Событие называется достоверным, если оно должно произойти при каждой реализации комплекса условий.
8. Два события A и B называется несовместимым, если их совместный исход невозможен.
9. События $B_1, B_2, \dots, B_n$ – образуют полную группу событий, если хотя бы одно из них непременно должно произойти при каждом осуществлении комплекса условий.
10. В каждой задаче теории вероятности приходится иметь дело с каким-либо определенным комплексом условий из какой-либо определенной системы S событий, наступающих или не наступающих.

Относительно системы S поставим некоторые допущения:

- Если системе S принадлежат события A, B, то ей принадлежат $A \cap B$, $A \cup B$, $A - B$ (замкнута относительно операций).
- Система S должна содержать достоверные и невозможные события. Если выполняется 1-7 то система называется полем событий.

Определение: Элементарные события – такие события в системе S , которые не могут быть разложены на составляющие события.

Пример № 11: Разложить на составляющие события: выпадение конкретного числа очков при бросании кубика.

Пусть заданно некоторое множество L . Элементы которого называются элементарными состояниями. Предположим, что фиксирована некоторая система α подмножеств, эти подмножества называются просто события, которая обозначается A, B, C, D и т.д. при этом потребуем:

- само множество L есть событие;
- если A событие, то $\bar{A}$ тоже событие;
- если A_1, A_2, A_3 – события, то и $A_1 + A_2, \dots, A_1 * A_2$ и $A_1 - A_2$ – тоже событие.

Тогда **аксиомы теории вероятности** можно сформулировать следующим образом:

- Каждому событию A поставлено в соответствие неотрицательное число $P(A)$ называемое вероятности события A ;
- Если событие A_1, A_2 – попарно совместны, то вероятность $P(A_1 + A_2 + A_3 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$;
- Вероятность множества – $L = 1. P(L) = 1$;
- Если A и B два случайных события, то $P\left(\frac{A}{B}\right) = \frac{P(A \cdot B)}{P(B)}$ -вероятность отношения A и B

называется условной вероятностью.

2.6 Принцип практической невозможности маловероятных событий

При решении многих практических задач приходится иметь дело с событиями, вероятность которых мала, то есть близка к нулю. Можно ли считать, что маловероятное событие A в единичном испытании не произойдет? Такое заключение сделать нельзя, так как не исключено, хотя и мало вероятно, что событие A наступит.

Казалось бы, появление или непоявление маловероятного события в единичном испытании предсказать невозможно. Однако длительный опыт показывает, что маловероятное событие в единичном испытании в подавляющем большинстве случаев не наступает. На основании этого факта принимают следующий «принцип практической невозможности маловероятных событий»: *если случайное событие имеет очень малую вероятность, то практически можно считать, что в единичном испытании это событие не наступит.*

Для каждой задачи маловероятное событие имеет свои определенные значения и сказать однозначно какой должна быть вероятность нельзя. Например, если вероятность, того что парашют при прыжке не раскроется, равна 0,01, то было бы недопустимым применять такие парашюты. Если же вероятность того, что поезд дальнего следования прибудет с опозданием, равна 0,01, то можно практически быть уверенным, что поезд прибудет вовремя.

Достаточно малую вероятность, при которой (в данной определенной задаче) событие можно считать практически невозможным, называют уровнем значимости. На практике обычно принимают уровни значимости, заключенные между 0,01 и 0,05. уровень значимости, равный 0,01 называют однопроцентным; уровень значимости равный 0,02, называют двухпроцентным и так далее.

Рассмотренный принцип позволяет делать предсказания не только о событиях, имеющих малую вероятность, но и о событиях, вероятность которых близка к единице. Действительно, если событие A имеет вероятность, близкую к нулю, то вероятность $\bar{A}$ близка к единице. С другой стороны, непоявление события A означает наступление противоположного события $\bar{A}$. Таким образом, из принципа невозможности маловероятных событий вытекает следующее важное следствие: *если случайное событие имеет вероятность, очень близкую к единице, то практически можно считать, что в единичном испытании это событие наступит.* Разумеется, и здесь ответ на вопрос о том, какую вероятность считать близкой к единице, зависит от сущности задачи.

Задания:

1. Бросаются две игральные кости. Найдите вероятности событий:

1) сумма выпавших очков равна 2, 3, ..., 12. Полученные результаты изобразите на координатной плоскости точками, у которых абсцисса равна сумме очков, а ордината – вероятности этой суммы;

2) разность выпавших очков равна 0, 1, ..., 5;

3) сумма выпавших очков больше их произведения;

2. В ящике 50 деталей, из них 5 окрашенных. Наудачу вынимают одну деталь. Найти вероятность того, что извлеченная деталь окажется окрашенной.

3. Из пяти отрезков длиной 1, 3, 5, 7 и 9 наудачу выбирают три. Какова вероятность того, из них можно построить треугольник?

4. Слово «молния» разрезали на буквы и выложили наудачу четыре из них в ряд. Какова вероятность получить слово «мили»?

5. В ящике лежат 20 одинаковых на ощупь шаров: 12 белых и 8 черных. Какова вероятность вынуть наудачу 1) белый шар; 2) черный шар; 3) два разноцветных шара?

6. В партии из 100 деталей отдел технического контроля обнаружил 5 нестандартных деталей. Чему равна частота появления нестандартных деталей?

7. На карточках выписаны цифры 1, 2, 3, 4, 5, 6, 7, 8 и 9. Наугад берут четыре карточки и выкладывают их в ряд. Какова вероятность того, что: 1) получится четное число; 2) получится число 1234?

8. Внутри круга радиуса r наудачу брошена точка. Найти вероятность того, что точка окажется внутри вписанного в круг квадрата.

9. Двое договорились о встрече на следующих условиях: каждый приходит в указанное место независимо друг от друга и наудачу в любой момент времени от 13.00 до 14.00. Придя, ожидает не более получаса, а уходит не позднее 14.00. Какова вероятность того, что они встретятся?

10. В упражнении 11 время ожидания сокращено до 20 мин. Какова вероятность встречи?

11. В окружность вписан квадрат. В круг, ограниченный этой окружностью, наудачу бросается точка. Какова вероятность, что она попадет в квадрат?

12. В окружность наудачу вписывается треугольник. Какова вероятность, что он: 1) прямоугольный; 2) равнобедренный; 3) тупоугольный?

Тест для самоконтроля

1. Установите соответствия:

1). Классическая вероятность это

а). отношение числа испытаний, в которых событие появилось к общему числу фактически произведенных испытаний;

2). Относительной частотой события называют...

б). отношение числа благоприятных исходов к числу всех исходов;

3). Произведением двух событий называют

в). событие которое происходит тогда и только тогда, когда имеют место оба события А и В

4). Суммой двух событий называют...

г). Событие состоящее в появлении события А или события В, или обоих этих событий

Ответьте на вопросы:

2. Чему равно $P(A) + P(\bar{A})$?

3. Чему равна вероятность достоверного события?

4. Игральная кость бросается два раза. Какова вероятность того, что сумма выпавших очков = 4? Выбери правильный ответ.

А). 0,5

б). 0,1

в). 0,83

5. Участники жеребьевки тянут из ящика жетоны с номерами от 1 до 100. Чему равна вероятность того, что номер первого наудачу извлеченного жетона не содержит цифры 5?
6. Укажите возможные исходы при бросании игральной кости:
7. Чему равна вероятность случайного события?
8. Что вероятнее выбросить при метании двух костей – 7 очков или 8 очков?

Тема №3 Условная вероятность. Формула Байеса. Схема Бернулли

Изучив эту тему, Вы будете:

Иметь представление:

- о теории вероятности как науке.

Знать:

- основные понятия: зависимые и независимые события. Условная вероятность, гипотеза;
- формулу полной вероятности, формулу Байеса;
- формулу Бернулли.

Уметь:

- определять вид событий;
- пользоваться основными понятиями, формулами теории вероятностей при решении задач.

Входной тест

- 1) Игральная кость бросается два раза. Какова вероятность того, что сумма выпавших очков = 6?
- 2) Является ли следующее событие достоверным: «Кит-это млекопитающие»
- 3) Запишите формулу для нахождения геометрической вероятности.
- 4) Чему равно $P(A) + P(\bar{B})$?
- 5) Сколько перестановок можно составить из букв: п, р, о, г, р, а, м, м, а.

3.1 Условная вероятность. Теорема полной вероятности

В теме №2 было определено случайное событие как событие, которое при осуществлении совокупности условий S может произойти или не произойти. Если при вычислении вероятности события никаких других ограничений, кроме условий S , не налагается, то такую вероятность называют безусловной. Если же налагаются и другие дополнительные условия, то вероятность события называют условной.

Пусть вероятность события B не зависит от появления события A . Тогда:

Определение: Событие A называется *независимым* от события B , если вероятность события A не зависит от того, наступило событие B или нет. В противном случае событие A называется *зависимым* от события B .

При совместном рассмотрении двух событий A и B возникает вопрос: На сколько связаны эти события друг с другом? В какой мере? Наступление одного из них влияет на возможное наступление другого? Давайте введем понятие условная вероятность.

Определение: Вероятность наступления события A при условии, что событие B наступило, называется *условной вероятностью* и обозначается: $P_B(A)$.

Формула умножения вероятностей.

Вероятность произведения события А и В равна произведению вероятности события А на условную вероятность события В при условии, что А произошло, т.е.

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

Следствие: если случайное событие В не зависит от А, то и А не зависит от В.

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

Пример 1: В урне находятся 3 черных и 2 белых шара. Из урны извлекают последовательно (без возвращения) два шара.

Решение.

А – [событие состоит в том, что первым будет изъят белый шар].

В – [событие состоит в том, что вторым будет изъят черный шар].

Найти вероятность произведения событий А и В (т.е. их совместного наступления). $P(A \cdot B) = P(A) \cdot P_A(B)$

$$P(A) = \frac{2}{5}$$

Извлекли 1 шар; Сколько осталось ? (4) (Зависимость)

$$P_A(B) = \frac{3}{4}$$

$$P(AB) = \frac{2}{5} \cdot \frac{3}{4} = \frac{6}{20} = \frac{3}{10};$$

Пример 2: Пусть имеется а белых и b черных мешков, причем в каждом белом мешке х красных и у синий шаров, а в каждом черном мешке – u красных и v синих шаров. Сначала случайным образом выбирают один мешок, а потом из него вынимают шар. Найдём вероятность P(B), P(Ч), P(B · К) и условные вероятности $P_B(K)$, $P_Ч(K)$.

Решение.

Очевидно, что вероятность того, что выбран белый мешок равна- $P(A) = \frac{a}{a+b}$, а

$$P(Ч) = \frac{b}{a+b}.$$

Далее условная вероятность того, что выбран красный шар из белого мешка равна -

$$P_B(K) = \frac{x}{x+y}, \text{ а из черного мешка } P_Ч(K) = \frac{u}{u+v}.$$

Событие B·K состоит в том, что выбран белый мешок, а из него извлечен красный шар. По формуле умножения вероятность этого события равна: $P(B \cdot K) = P(B) \cdot P_B(K) = \frac{a}{a+b} \cdot \frac{x}{x+y}.$

Задание 1 Аналогично определите: $P(Ч \cdot K) =$ _____

$P(B \cdot C) =$ _____ $P(Ч \cdot C) =$ _____

Задание 2: Определите и запишите вероятность выпадения двух гербов при двукратном бросании.

А- [событие состоящее в том, что при первом бросании монеты выпадет герб].

В- [событие состоящее в том, что при втором бросании монеты выпадет герб].

Тогда $P(A \cdot B) =$ _____

Вероятность события А, которое может наступить лишь при условии появления одного из несовместимых событий $B_1, B_2, \dots, B_n$ образующих полную группу, равна сумме произведений вероятностей каждого из этих событий на соответствующую условную вероятность события А.

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

Алгоритм решение задач пункта 3.1

1. Определить события искомой вероятности.
2. Определить вид событий.
3. Исходя из вопроса задачи и вида событий найти в таблице 1 искомую формулу.
4. подставить исходные данные в формулу, подсчитать и записать ответ.

Таблица №1

События	Совместные	Несовместные
Зависимые	$P(A + B) = P(A) + P(B) - P(A \cdot B)$ $P(A \cdot B) = P(A) \cdot P_A(B)$	$P(A + B) = P(A) + P_A(B)$ $P(A \cdot B)$ - нет
Независимые	$P(A + B) = P(A) + P(B) - P(AB)$ $P(A \cdot B) = P(A) \cdot P(B)$	$P(A + B) = P(A) + P(B)$ $P(A \cdot B)$ - нет

Приглашение

Используя алгоритм решим следующие задачи.

Задача 1. Два спортсмена независимо друг от друга стреляют по одной мишени. Вероятность попадания первого спортсмена = 0,7; а второго = 0,8. Какова вероятность того, что мишень будет поражена?

Решение.

В мишень попадет либо первый, либо второй, либо попадут вместе- **хотя бы** попадет один стрелок, т.е. произойдет $A + B$.

A и B – независимы, совместны. Из таблицы №1

$P(A + B) = P(A) + P(B) - P(AB)$. Подставим исходные данные, получим:

$$P(A + B) = 0,7 + 0,8 - 0,7 \cdot 0,8 = 0,94;$$

Задача 2. Бросается игральный кубик. Какова вероятность того, что выпадет четное число и число меньше 5?

Решение: Что будем понимать под событием A ?

A – [событие появления четного числа очков 2, 4, 6].

$$P(A) = \frac{3}{6}$$

B – [появление числа очков < 5].

Благоприятные исходы: 1, 2, 3, 4.

$$P(B) = \frac{2}{4} = \frac{1}{2}$$

A и B – независимы и совместны.

$$P(A \cdot B) = \frac{3}{6} \cdot \frac{1}{2} = 0,25$$

Задача 3.

Пусть вероятность того, что студент опоздает на лекцию, равна 0,08. Найти вероятность того, что среди 96 студентов на лекцию опоздает не более трех человек.

Решение.

Решение задачи основывается на вычислении вероятностей $P_n(k)$.

Событие A – «на лекцию опоздает не более 3-х человек» означает, что опоздает или 0, или 1, или 2, или 3 студента, т. е. $k = 0$, или $k = 1$, или $k = 2$, или $k = 3$.

Искомая вероятность определяется:

$$P(A) = P_{96}(0) + P_{96}(1) + P_{96}(2) + P_{96}(3).$$

Вычислим вероятности $P_n(k)$ разными способами.

1) по биномиальной формуле (формуле Бернулли)

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

где $C_n^k = \frac{n!}{k! \cdot (n-k)!}$, $p = 0,08$; $q = 1 - p = 0,92$; $n = 96$.

$$P_{96}(0) = 0,92^{96} = 0,00033;$$

$$P_{96}(1) = 96 \cdot 0,08 \cdot 0,92^{95} = 0,00279;$$

$$P_{96}(2) = \frac{96!}{2! \cdot 94!} \cdot 0,08^2 \cdot 0,92^{94} = 0,01151;$$

$$P_{96}(3) = \frac{96!}{3! \cdot 93!} \cdot 0,08^3 \cdot 0,92^{93} = 0,03137.$$

$$P(A) \approx 0,00033 + 0,00279 + 0,01151 + 0,03137 = 0,04600.$$

Задача 4.

Имеется два набора деталей. Вероятность того, что деталь первого набора стандартна = 0,8, а второго – 0,9. Найти вероятность того, что взятая наудачу деталь (из наудачу взятого набора) – стандартна.

Решение:

A – [событие, что извлеченная деталь стандартна];

B_2 – [извлечена из второго набора];

B_1 – [извлечена из первого набора];

Тогда

$$P(B_1) = \frac{1}{2} \quad P_{B_1}(A) = 0,8$$

$$P(B_2) = \frac{1}{2} \quad P_{B_2}(A) = 0,9$$

По формуле полной вероятности имеем

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) = \frac{1}{2} \cdot 0,8 + \frac{1}{2} \cdot 0,9 = 0,85; \text{ Вероятность того, что взятая наудачу деталь стандартна } 0,85.$$

3.2 Формула Байеса

Пусть A может наступить при условии появления одного из несовместных событий $B_1, B_2, \dots, B_n$ образующие полную группу. Поскольку заранее неизвестно; какое из них наступит, их называют *гипотезами*. Вероятность появления события A определяется по формуле:

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + \dots + P(B_n) \cdot P_{B_n}(A) \quad (1)$$

Пусть произведено испытание в результате которого появилось событие A .

- Выясним, как изменились вероятности гипотез? *Другими словами найдем*

$$P_A(B_1), P_A(B_2), \dots, P_A(B_n)$$

Найдем $P_A(B_1)$ по теореме умножения.

$$P(AB_1) = P(A) \cdot P_A(B_1) = P(B_1) \cdot P_{B_1}(A)$$

$$\text{Выразим } P_A(B_1): P_A(B_1) = \frac{P(B_1) \cdot P_{B_1}(A)}{P(A)}$$

Заменим $P(A)$ на (1) получим

$$P_A(B_1) = \frac{P(B_1) \cdot P_{B_1}(A)}{P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + \dots + P(B_n) \cdot P_{B_n}(A)}$$

аналогично выводятся формулы для определения других гипотез. Обозначим: B_i ($i = 1, 2, \dots, n$).

$$P_A(B_i) = \frac{P(B_i) \cdot P_{B_i}(A)}{P(B_1) \cdot P_{B_1}(A) + \dots + P(B_n) \cdot P_{B_n}(A)} \quad \text{Формула Байеса}$$

Вывод: Формула Байеса позволяет переоценить вероятности гипотез после того, как становится известным результат испытания, в итоге которого появилось событие A .

Пример 3: Детали изготавливаемые цехом завода, попадают для проверки их на стандартность к одному из двух контролеров. Вероятность того, что деталь попадет к первому 0,6; ко

второму = 0,4. Вероятность того, что годная деталь будет признана стандартной первым контролером = 0,94; вторым = 0,98. Годная деталь при проверке оказалась стандартной. Найти вероятность того, что эту деталь проверил первый контролер.

Решение.

A – [событие, состоящее в том, что годная деталь признана стандартной].

Какие гипотезы можно выделить?

а) проверил первый контролер (B_1)

б) проверил второй контролер (B_2)

Искомую вероятность найдем по формуле Байеса $P_A(B_1) = \frac{P(B_1) \cdot P_{B_1}(A)}{P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A)}$

$$P(B_1) = 0,6 \quad P_{B_1}(A) = 0,94$$

$$P(B_2) = 0,4 \quad P_{B_2}(A) = 0,98.$$

$$P_A(B_1) = \frac{0,6 \cdot 0,94}{0,6 \cdot 0,94 + 0,4 \cdot 0,98} = 0,5881$$

До испытания вероятность гипотезы = 0,6, а после того как стал известен результат испытания $P(B_1) = 0,59$ (более точнее).

3.3 Схема Бернулли. Формула Бернулли

При введении вероятности отмечалось, что если вероятность некоторого события A равна p , то вероятнее всего, что при повторен испытаний много раз относительная частота благоприятных этому событию исходов будет мало отличаться от значения p . Это утверждение, называемое в теории вероятностей *законом больших чисел*, лежит в основе всех практических приложений этой теории- оно позволяет с помощью вычисленных вероятностей предсказать частоту наступления данного события в длинной серии независимых испытаний. Если производится несколько испытаний, причем вероятность события A в каждом испытании не зависит от исходов других испытаний, то такие испытания называются *независимыми относительно события A* .

Пусть производится n независимых испытаний, в каждом из которых событие A может появиться либо не появиться. Условимся считать, что вероятность события A в каждом испытании постоянна и равна p . Следовательно, вероятность ненаступления события A в каждом испытании так же постоянна и равна $q=1-p$.

Введем формулу Бернулли, позволяющую вычислить вероятность того, что в серии из n независимых испытаний событие A , имеющее вероятность p , встретится m раз.

Найдем вероятность того, что в результате проведенных n испытаний событие A произошло m раз. $P_n(m) - ?$

Рассмотрим композицию n независимых испытаний и построим композиционное пространство элементарных событий.

Общий вид элемента этого пространства следующий:

$$A^{\alpha_1} \cdot A^{\alpha_2} A^{\alpha_3} \dots A^{\alpha_n}, \text{ где } A^1 = A, \alpha_i = \{0,1\}.$$

При этом вероятность наступления такого события равна:

$$p^{\sum_{i=1}^n \alpha_i} \cdot q^{n - \sum_{i=1}^n \alpha_i} \quad (\text{умножение при независимых событиях}).$$

Найдем вероятность наступления любого элементарного события из композиционного пространства:

$$P(A^{\alpha_1} \cdot A^{\alpha_2} \dots A^{\alpha_n}) = \prod_{j=1}^n P(A^{\alpha_j}) = p^m \cdot q^{n-m}$$

Рассмотрим в композиционном вероятностном пространстве событие: в n испытаниях событие A произошло m раз. Событие A состоит из C_n^m - общее кол-во элементарных событий, в которое входит событие A .

А произошло m раз, $\bar{A}$ - $n-m$ раз. Вероятность каждого из этих элементарных событий одинакова и равна: $p^n \cdot q^{n-m}$

Следовательно результат равняется:

$$P_n(m) = C_n^m \cdot p^m \cdot q^{n-m}, \text{ где } C_n^m = \frac{n!}{m!(n-m)!}.$$

Полученную формулу называют **формулой Бернулли**.

Пример 4: Монета подбрасывается 10 раз. Какова вероятность того, что орел выпадает при этом 3 раза.

Решение.

$A = [\text{событие состоящие в том, что орел выпадет при 1 бросании монеты}]$.

$$P(A) = \frac{1}{2}, \quad n = 10, \quad m = 3, \quad q = 1 - P(A) = 1 - \frac{1}{2} = \frac{1}{2}$$

тогда по формуле Бернулли получаем:

$$P_{10}(3) = C_{10}^3 \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^{10-3} = \frac{10!}{3! \cdot 7!} \cdot \frac{1}{2^3} \cdot \frac{1}{2^7} = \frac{10 \cdot 9 \cdot 8}{6 \cdot 2^3 \cdot 2^7} = 0,1$$

Задание 4: Найти вероятность того, что в ближайшие 6 суток расход электроэнергии в течении 4 суток не превысит нормы. Вероятность того, что расход- электроэнергии- в продолжении одних суток не превысит установленной нормы, равна 0,75.

$p = \underline{\hspace{1cm}}; q = \underline{\hspace{1cm}}; m = \underline{\hspace{1cm}}; n = \underline{\hspace{1cm}};$

$P_{\underline{\hspace{1cm}}}(\underline{\hspace{1cm}}) = \underline{\hspace{2cm}}$

Задачи.

1. Два стрелка произвели по одному выстрелу. Вероятность попадания в мишень первым равна 0,7, а вторым – 0,6. Найти вероятность того, что хотя бы один из стрелков попал в мишень.
2. Бросается игральный кубик. Какова вероятность того, что выпадет четное и меньшее 5 число очков?
3. В ящике лежат 12 красных, 8 зеленых и 10 синих шаров, одинаковых на ощупь. Наудачу вынимается один шар. Какова вероятность того, что он красный, если известно, что он не синий?
4. В первом ящике содержится 20 деталей из них 15 стандартных. Во втором – 30 деталей из них 24 стандартных. В третьем – 10 деталей из них 6 стандартных. Найти вероятность того, что наудачу извлеченная деталь из наудачу взятого ящика стандартная.
5. В группе спортсменов 20 биатлонистов, 6 велосипедистов и 4 бегуна. Вероятность выполнить квалификационную норму такова: у биатлонистов-0,9, у велосипедистов-0,8 и у бегунов – 0,75. Найти вероятность того, что спортсмен выбранный наудачу, выполнит норму.
6. Партия электрических лампочек на 20 % изготовлена заводом 1, на 30 % - заводом 2 и на 50 % - заводом 3. Для завода 1 вероятность выпуска бракованной лампочки равна 0,01, для завода 2- 0,005 и для завода 3- 0,006. Какова вероятность того, что взятая на удачу из партии лампочка оказалась бракованной?
7. В трех залах кинотеатра идут три различных фильма. Вероятность того, что на определенный час в кассе 1-го зала есть билеты, равна 0,3, в кассе 2-го зала – 0,2, в кассе 3-го зала – 0,4. Какова вероятность того. Что на данный час имеется возможность купить билет хотя бы на один фильм?
8. В цехе 6 моторов. Для каждого мотора вероятность того, что он в данный момент включен, равна 0,8. найти вероятность того, что в данный момент: а). включены 4 мотора; б). включены все моторы; в). выключены все моторы.

Тест для самоконтроля

1. Продолжите предложение. Событие А называется зависимым от события В если ...
2. Определите правильный ответ. В группе спортсменов 20 лыжников, 6 конькобежцев и 4 биатлониста. Вероятность выполнить квалификационную норму такова: лыжники- 0,9; конькобежцы- 0,8; биатлонисты- 0,75. Вероятность того, что спортсмен выбранный наудачу, выполнит норму равна:
а). 0,98
б). 0,86
в). 1

Ответьте на вопросы:

3. Какую формулу необходимо использовать в задачи для нахождения вероятности: На фабрике изготавливающей шурупы, первый станок производит 35%, второй 65% всех изделий. Брак в их продукции составляет 5% и 4%. Какова вероятность того, что случайно выбранный шуруп окажется дефектным?
4. Определите правильный ответ. В двух ящиках имеются радиолампы. В первом ящике содержится 12 ламп, из них 1 нестандартная; во втором 10 ламп, из них 1 нестандартная. Из первого ящика наудачу взята лампа и переложена во второй. Найти вероятность того, что наудачу извлеченная из второго ящика лампа будет нестандартной.
А). $\frac{13}{132}$
б). $\frac{27}{124}$
в). $\frac{17}{223}$
5. Вероятность поломки первого станка в течении смены равна 0,2, а второго- 0,13. Чему равна вероятность того, что оба станка потребуют наладки в течении смены?
6. Запишите формулу вероятности того, что событие А появиться в пяти независимых испытаниях не менее двух раз, если в каждом испытании вероятность появления события А равна р.

РАЗДЕЛ №2 СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Цели изучения: расширить и систематизировать сведения о событии, вероятности. Сформировать навыки решения задач на определения закона распределения, функции распределения, плотности распределения, вида распределения, числовых характеристик случайной величины.

Тема №4 Случайная величина

Изучив эту тему, Вы будете:

Иметь представление:

- о случайных величинах; видах случайных величин; функции распределения; числовых характеристиках случайных величин; о законах распределения случайной величины;

Знать:

- способы задания случайной величины;
- способы представления и преобразования функции распределения;
- способ нахождения функции распределения, плотности распределения;
- числовые характеристики случайных величин;

Уметь:

- по заданному закону распределения определять с помощью STADIA, математического пакета Mathcad, ТП Excel:
 - функцию распределения;
 - плотность распределения;

- определять числовые характеристики случайных величин;
- строить статистические графики и таблицы;
- применять различные виды распределений для нахождения вероятности случайных величин, композиции законов распределения.

Входной тест

1. Ответьте на вопрос. Чему равна вероятность совместного появления двух зависимых события?
2. Запишите формулу для нахождения статистической вероятности.
3. Множество действительных чисел обозначается буквой...
 - а). R; б). S;
 - в). L; г). Z.
4. Выберите из перечисленного правильный способ выбора столбца из заданной матрицы $A = \begin{pmatrix} 1 & 0,4 \\ 2 & 0,6 \end{pmatrix}$ в МП MathCAD:
 - а). $A^{(1)}$
 - б). $A^{(0)}$
 - в). $T = A(1)$
 - г). $T = A_0$
5. Укажите результат интегрирования функции $f = \int_1^2 dx$ P=1
6. Установите соответствия горячих клавиш и действий производимых в MathCAD:

1. Shift+ F9	а) вставка матрица
2. Ctrl+M	б) расчет неопределенного интеграла
3. F9	в) построение графика в ручном режиме вычислений

4.1 Случайная величина

Результаты случайных экспериментов, как правило, выражаются числами. Одной из типичных причин для этого является использование измерительных приборов с числовыми шкалами. В этом случае результаты случайных экспериментов принято называть случайными величинами.

Определение. Случайной величиной называется измеримая числовая функция x , элементами которой являются элементарные события, которая в результате опыта, может принять всевозможные значения из множества допустимых значений.

Если произведено испытание, в результате которого произошло некоторое элементарное событие x_i , то в соответствии с функцией x этому элементарному событию соответствует число, которое называется реализацией случайной величины x в данном испытании.

Пример №1

Опыт – бросание игральной кости; случайная величина X -число выпавших очков. Множество возможных значений $\{1,2,3,4,5,6\}$

Пример №2

Опыт- работа ЭВМ после очередного ремонта; случайная величина T -время наработки ЭВМ до первого отказа (сбоя). Множество возможных значений несчетно.

Определение. Случайная величина называется дискретной, если в результате испытания она может принять значение из конечного либо счетного множества возможных числовых значений.

Определение. Случайная величина называется непрерывной, если в результате испытания она может принять значение из отрезка, интервала или полуинтервала либо множества возможных числовых значений.

Пример №3

Примеры непрерывных случайных величин: возраст студентов, длина ступни ноги человека, масса детали и т. д.

Это положение относится ко всем случайным величинам, измеряемым на непрерывной шкале, таким как меры веса, длины, времени, температуры, расстояния. Измерение может быть проведено с точностью до какого-нибудь десятичного знака, но случайная величина – теоретически непрерывная величина. В экономическом анализе находят широкое применение относительные величины, различные индексы экономического состояния, которые также вычисляются с определенной точностью, скажем, до двух знаков после запятой, хотя теоретически их значения – непрерывные случайные величины.

У непрерывной случайной величины возможные значения заполняют некоторый интервал (или сегмент) с конечными или бесконечными границами.

Случайные величины в дальнейшем будем обозначать большими буквами: X, Y, Z

Совокупность значений может быть задана таблицей, функцией или графиком.

Соотношение, устанавливающее связь между отдельными возможными значениями случайной величины и соответствующими им вероятностями, называется **законом распределения дискретной случайной величины**.

Простейшей формой закона распределения для дискретных случайных величин является ряд распределений.

Рядом распределения дискретной случайной величины X называется таблица, в которой перечислены возможные (различные) значения этой случайной величины $x_1, x_2, \dots, x_n$ с соответствующими им вероятностями $p_1, p_2, \dots, p_n$.

x_i	x_1	x_2	$\dots$	x_n
p_i	p_1	p_2	$\dots$	p_n

Таким образом, случайная величина X в результате испытания может принять одно из возможных значений $x_1, x_2, \dots, x_n$ с вероятностями

$$P(X = x_1) = p_1; P(X = x_2) = p_2; P(X = x_n) = p_n.$$

Можно использовать более короткую запись: $P(x) = P(5) = 0,2$. Так как события $(X = x_1), (X = x_2), \dots, (X = x_n)$ составляют полную группу событий, то сумма вероятностей $p_1, p_2, \dots, p_n$ равна единице:

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

Ряд распределения случайной дискретной величины должен удовлетворять следующим условиям:

$$\begin{aligned} P(x) &\geq 0, \\ \sum_{i=1}^n P(x_i) &= 1. \end{aligned} \quad (3.1)$$

Пример №4 Рассматривается работа трех независимых работающих технических устройств (ТУ); вероятность нормальной работы первого ТУ равна 0,2, второго – 0,4, третьего – 0,5; Построить ряд распределения случайной величины X .

Решение:

Случайная величина X – число работающих ТУ. Возможные значения случайной величины: 0, 1, 2, 3.

Определим соответствующие вероятности, пользуясь правилами сложения и умножения. Для краткости будем обозначать нормальную работу знаком “+”, а отказ – знаком “-”.

$$p_1 = P\{X = 0\} = P\{- - -\} = 0,8 \cdot 0,6 \cdot 0,5 = 0,24$$

$$\begin{aligned} p_2 = P\{X = 1\} &= P\{+ - -\} + P\{- + -\} + P\{- - +\} = 0,2 \cdot 0,6 \cdot 0,5 + \\ &+ 0,8 \cdot 0,4 \cdot 0,5 + 0,8 \cdot 0,6 \cdot 0,5 = 0,46 \end{aligned}$$

$$p_1 = P\{X = 2\} = P\{- + +\} + P\{+ - +\} + P\{+ + -\} = 0,8 \cdot 0,4 \cdot 0,5 + 0,2 \cdot 0,6 \cdot 0,5 + 0,2 \cdot 0,4 \cdot 0,5 = 0,26$$

$$p_1 = P\{X = 3\} = P\{+ + +\} = 0,2 \cdot 0,4 \cdot 0,5 = 0,04$$

Ряд распределения случайной величины X имеет вид:

X	0	1	2	3
p	0,24	0,46	0,26	0,04

Задание №1

Постройте ряд распределения случайной величины X , для следующей задачи:

В книжном магазине организована лотерея. Разыгрываются две книги стоимостью по 10 руб. и одна – стоимостью в 30 руб. Составить закон распределения случайной величины X – суммы чистого (возможно-го) выигрыша для того, кто приобрел один билет за 1 руб., если всего продано 50 билетов.

x_i			
$P(x_i) = p_i$			

4.2 Функция распределения

При анализе экономических явлений определенный смысл имеют *кумулятивные* (накопленные) вероятности случайных величин. Нас может интересовать вероятность того, что число проданных единиц некоторого товара окажется не меньше некоторого определенного числа, гарантирующего прибыль продавцу, вероятность того, что суммы возможных убытков от рискованных инвестиций окажутся не выше (или только меньше) некоторого определенного значения, и т. д. Зная закон распределения дискретной случайной величины, можно составить функцию накопленных вероятностей. Определим **интегральную (кумулятивную) функцию распределения**.

Функцией распределения дискретной случайной величины называется функция $F(x)$, определяющая для каждого значения x вероятность того, что случайная величина X не превзойдет некоторого x , т. е.

$$F(x) = P(X \leq x) = \sum_{\substack{i \\ x_i \leq x}} P(X = x_i) = \sum_{\substack{i \\ x_i \leq x}} P(X = i), \quad (3.2)$$

где суммирование распространяется на все значения индекса i , для которых $x_i \leq x$.

Функцию $F(x)$ называют также накопленным (кумулятивным) распределением вероятностей. Иногда вместо термина «**функция распределения**» используют термин «**интегральная функция распределения**».

Свойства:

1. $F(-\infty) = 0$
 $F(+\infty) = 1$ или $0 \leq F(x) \leq 1$;
2. $F(t)$ - монотонна не убывающая;
3. $P(t_1 < x < t_2) = F(t_2) - F(t_1)$.

Построим функцию распределения для дискретной случайной величины. Для удобства договоримся, что случайные величины располагаются в порядке возрастания.

$$X = \begin{Bmatrix} x_1 & x_2 & \dots & x_s \\ p_1 & p_2 & \dots & p_s \end{Bmatrix}$$

$$F(x) = P(X \leq x)$$

т.е. по определению для любого действительного X , $F(x)$ численно равно вероятности наступления следующего события: в результате испытаний над X оно приняло значение строго меньше x .

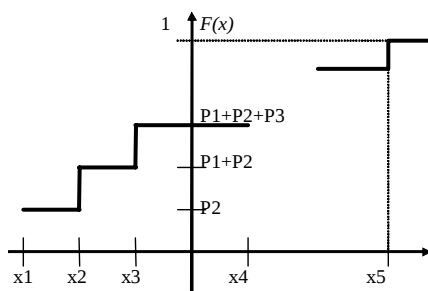


Рис.4 функция распределения для дискретной случайной величины

Закон распределения непрерывной случайной величины можно задать в виде **интегральной функции распределения**, являющейся наиболее общей формой задания закона распределения случайной величины

Функция распределения (или интегральная функция) $F(x)$ – универсальная форма задания закона распределения случайной величины. Для непрерывной случайной величины функция распределения также определяет вероятность того, что случайная величина X примет значение, меньшее фиксированного действительного числа x , т. е. $F(x) = P(X < x)$.

При изменении x меняются вероятности $P(X < x) = F(x)$. Поэтому $F(x)$ и рассматривают как функцию переменной величины. Принято считать, что случайная величина X известна, если известна ее функция распределения $F(x)$.

Теперь можно дать более точное определение непрерывной случайной величины: **случайную величину называют непрерывной, если ее функция распределения есть непрерывная, кусочно-дифференцируемая функция с непрерывной производной.**

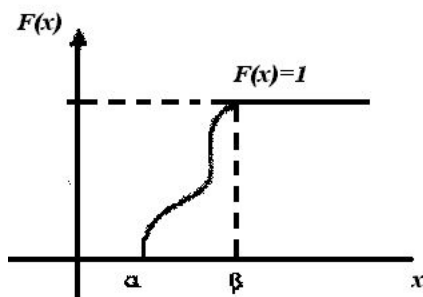


Рис.5.1. График функции распределения непрерывной случайной величины

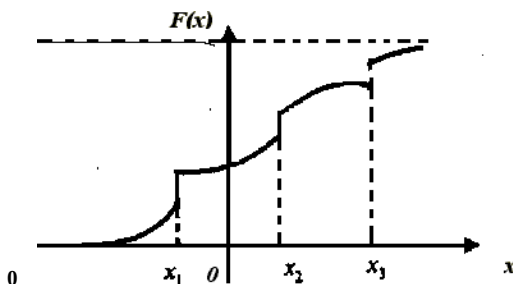
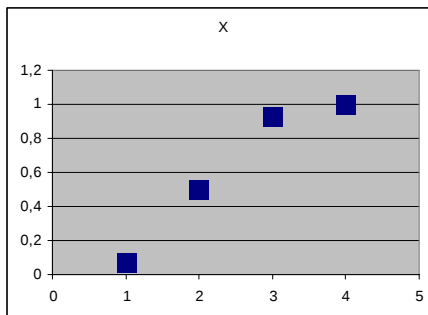


Рис. 5.2. График кусочно-непрерывной функции распределения

Пример №1: Задан ряд распределения. Построить функцию распределения.

X	1	2	3	4
P	1/14	6/14	6/14	1/14



при $x \leq 1$ $F(x) = P(x < 1) = 0$;

при $1 < x \leq 2$

при $2 < x \leq 3$

$$F(x) = P(x < 2) = \frac{1}{14};$$

$$F(x) = P(x < 3) = \frac{1}{14} + \frac{6}{14} = \frac{7}{14};$$

при $3 < x \leq 4$ $F(x) = P(x < 4) = \frac{1}{14} + \frac{6}{14} + \frac{6}{14} = \frac{13}{14}$

Рис.6 График функции распределения

при $x > 4$ $F(x) = \frac{1}{14} + \frac{6}{14} + \frac{6}{14} + \frac{1}{14} = 1$

Из рисунка видно, что $F(x)$ имеет 4 скачка по числу X . Если увеличить X , то число скачков так же становится больше, а сами скачки меньше, вследствие чего ступенчатая кривая становится более плавной. В этом случае дискретная случайная величина постепенно приближается к непрерывной.

Пример №2: Функция распределения случайной величины x задана выражением:

$$F(x) = \begin{cases} 0, & x < -\frac{\pi}{4} \\ a \sin(x - \frac{\pi}{4}) + \frac{1}{2}, & -\frac{\pi}{4} \leq x \leq \frac{3\pi}{4} \\ 1, & x \geq \frac{3\pi}{4} \end{cases}, \text{ Определить коэффициент } a?$$

Вероятность попадания X в интервал при $x = \frac{3\pi}{4}$ функция $F(x)$ равна 1, т. е.

$$a \sin(x - \frac{\pi}{4}) + \frac{1}{2} = 1, \text{ или } a \sin(\frac{\pi}{2}) + \frac{1}{2} = 1 \Rightarrow a = \frac{1}{2}.$$

Задание непрерывной случайной величины с помощью функции распределения не является единственным. Её можно задать с помощью другой функции, которая называется дифференциальной функцией распределения или плотностью распределения $f(x)$, которая используется для описания распределения вероятностей только непрерывной случайной величины.

Определение. Плотностью распределения называется первая производная интегральной функции распределения $F(x)$.

Свойства:

1. $\forall x, f(x) \geq 0$

2. $P(t_1 < x < t_2) = \int_{t_1}^{t_2} f(t) dt$; Геометрически этот результат равен площади криволинейной трапеции, ограниченной осью OX , кривой распределения $P(x)$ и прямыми $x = t_1$ $x = t_2$

$$3. \int_{-\infty}^{+\infty} f(t) dt = 1$$

$$4. F'(x) = f(x) \text{ или } F(x) = \int_{-\infty}^x f(t) dt$$

Пример №3: Задана плотность распределения, определить функцию распределения.

$$f(x) = \begin{cases} 0 & x < 0; \\ (x-3)^2/9, & \text{при } 0 \leq x \leq 3; \\ 0 & x > 3. \end{cases}$$

$$F(x) = \int_{-\infty}^0 0 dt + \int_0^x \frac{(t-3)^2}{9} dt + \int_x^{+\infty} 0 dt = 0 + \frac{(t-3)^3}{27} \Big|_0^x + 0 = \frac{(x-3)^3}{27} - \frac{(0-3)^3}{27} = \frac{(x-3)^3}{27} + 1;$$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{(x-3)^3}{27} + 1 & 0 \leq x \leq 3 \\ 0 & x > 3 \end{cases}$$

ПРАКТИЧЕСКАЯ РАБОТА №1

ДИСКРЕТНАЯ И НЕПРЕРЫВНАЯ СЛУЧАЙНАЯ ВЕЛИЧИНА

Цель работы: изучение функции и плотности распределения случайной величины.

Порядок выполнения работы:

Выполните работу в математическом пакете MathCAD.

Для этого:

1. Составьте ряд распределения дискретной случайной величины согласно индивидуальному заданию (см. таблицу №1)

Таблица 1

№	Опыт	Случайная величина
1	Производится 3 независимых опыта, в каждом из которых событие А появляется с вероятностью 0,4	Число появления события А в трех опытах
2	Бросание пары костей	Модуль разности выпавших очков
3	Производится 4 независимых опыта, в каждом из которых событие А появляется с вероятностью 0,3	Число появления события А в четырех опытах
4	Бросание монеты 10 раз	Число выпавших гербов

2. По полученному ряду постройте многоугольник распределения.
3. Определите функцию распределения.
4. Постройте график распределения дискретной случайной величины.
5. Решите следующие задачи:

Задача1.

Дана плотность распределения случайной величины. Найти параметр γ , функцию распределения, вероятность выполнения неравенства $2 < x < 2,6$.

$$f(x) = \begin{cases} \frac{1}{\gamma - 1,5} & \text{при } x \in [1,5; 3] \\ 0 & \text{при } x \notin [1,5; 3] \end{cases}$$

Задача2.

Задана случайная величина функцией распределения. Найти $f(x)$ и вероятность того, что в результате испытания случайная величина примет значение в $(0, \pi/4)$

$$F(x) = \begin{cases} 0 & \text{при } x \in [0; -\infty) \\ \cos^2(x) & \text{при } x \in (0; \pi] \\ 0 & \text{при } x \in [\pi; +\infty] \end{cases}$$

Задача3.

Число иногородних судов, прибывающих ежедневно под погрузку в определенный порт – случайная величина x , заданная так:

x_i	0	1	2	3	4	5
$P(X)=p_i$	0,1	0,2	0	0,1	0,1	0,1

- Убедиться, что задан ряд распределения.
- Найти функцию распределения.
- Используя функцию распределения, найдите вероятность того, что в заданный день прибудет от 1 до 4 грузовых судов (включая 1 и 4);
- Если в заданный день прибывает больше трех судов, то порт берет на себя ответственность за издержки вследствие необходимости нанимать дополнительных водителей и грузчиков. Чему равна вероятность того, что порт понесет дополнительные расходы в заданный день?

Образец выполнения:

Распределение дискретной случайной величины

ORIGIN := 1

$$A := \begin{pmatrix} -2 & 0 & 1 & 4 & 7 \\ 0.2 & 0.5 & 0.1 & 0.1 & 0.1 \end{pmatrix}$$

Распределение случайной величины сохранено в матрице A . $A_{1,i}$ — значения случайной величины, $A_{2,i}$ — соответствующие вероятности, $i = 1, 2, 3, 4, 5$.

Функция распределения случайной величины

$$F(x) := \begin{cases} 0 & \text{if } -\infty < x < A_{1,1} \\ A_{2,1} & \text{if } A_{1,1} \leq x < A_{1,2} \\ (A_{2,1} + A_{2,2}) & \text{if } A_{1,2} \leq x < A_{1,3} \\ (A_{2,1} + A_{2,2} + A_{2,3}) & \text{if } A_{1,3} \leq x < A_{1,4} \\ (A_{2,1} + A_{2,2} + A_{2,3} + A_{2,4}) & \text{if } A_{1,4} \leq x < A_{1,5} \\ 1 & \text{if } A_{1,5} \leq x < \infty \end{cases}$$

+

Разверните панель программных элементов щелчком по кнопке  и панель знаков отношений — щелчком по кнопке , и не закрывайте их, пока не закончите определение функции. Введите имя функции переменной x и знак присваивания, щелкните в панели программных элементов по кнопке  введите в помеченной позиции нуль, щелкните по кнопке  и введите неравенства, определяющие первый интервал изменения аргумента (символ ∞ можно ввести щелчком по соответствующей кнопке в панели ), затем перейдите во вторую строку определения функции, введите $A_{2,1}$ — имя переменной, содержащей значение P_1 , или число 0.2 — значение P_1 , выделите, нажав клавишу $\langle \text{Space} \rangle$, выражение для функции, щелкните по кнопке , введите неравенства, определяющие второй интервал изменения аргумента (знак $\leq$ можно ввести щелчком по соответствующей кнопке в панели отношений), выделите, нажав клавишу $\langle \text{Space} \rangle$, вторую строку определения функции, щелкните по кнопке  введите, действуя как выше, определение функции на следующем интервале.

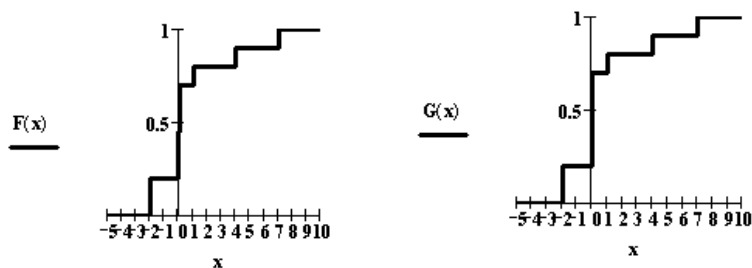
Функция распределения случайной величины, определенная другим способом

$$G(x) := \begin{cases} 0 & \text{if } -\infty < x < -2 \\ 0.2 & \text{if } -2 \leq x < 0 \\ 0.2 + 0.5 & \text{if } 0 \leq x < 1 \\ 0.2 + 0.5 + 0.1 & \text{if } 1 \leq x < 4 \\ 0.2 + 0.5 + 0.1 + 0.1 & \text{if } 4 \leq x < 7 \\ 1 & \text{if } 7 \leq x < \infty \end{cases}$$

Приведены два способа определения функции распределения: с использованием имен переменных и с использованием конкретных значений этих переменных.

График функции распределения случайной величины

Графики функций построены стандартным для декартовых графиков способом. Следует помнить, что Mathcad не совсем корректно строит графики ступенчатых функций, соединяя стрелками прямых значения функции в точке скачка. Более точный график функции распределения представляет собой отрезки, параллельные оси абсцисс, с "выкоп отым" правым концом.



4.2 ВЕРОЯТНОСТНЫЕ ХАРАКТЕРИСТИКИ ДИСКРЕТНОЙ И НЕПРЕРЫВНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Рассмотрим основные характеристики дискретной случайной величины при конечном числе значений.

Каждому значению дискретной случайной величины отвечает его вероятность. Как отмечалось выше, последовательность таких пар образует ряд распределения дискретной случайной величины:

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{bmatrix},$$

где $x_1 < x_2 < \dots < x_n$, $0 \leq p_i \leq 1$, $i = 1, \dots, n$, $\sum_{i=1}^n p_i = 1$.

Если случайная дискретная величина является случайной альтернативной величиной, т. е. задается двумя значениями 0 и 1 и соответствующими им вероятностями исходов $q = 1 - p$ и p , то ряд распределения принимает форму:

$$X = \begin{bmatrix} 0 & 1 \\ q & p \end{bmatrix},$$

где $0 \leq p \leq 1, p + q = 1$.

На основе ряда распределения можно определить среднее значение случайной дискретной величины как меру, которая объединяет значения случайной дискретной величины и их вероятности. **Среднее значение** есть взвешенная средняя всех возможных значений случайной величины, роль весов (частот) играют вероятности.

1. Ожидаемое среднее значение случайной величины называется **математическим ожиданием** $M(X)$. Математическим ожиданием дискретной случайной величины x с законом распределения:

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

(1)

называется число $M[x] = x_1p_1 + x_2p_2 + \dots + x_np_n$ при условии абсолютной сходимости ряда. Если x – непрерывная случайная величина с плотностью вероятности $f(x)$, то

$$M[x] = \int_{-\infty}^{\infty} xf(x) dx.$$

Свойства математического ожидания:

1. $M[c] = c$, если c постоянная.
2. $M[cx] = c M[x]$, если c постоянная.
3. $M[x_1 + \dots + x_n] = M[x_1] + \dots + M[x_n]$.

Если $x_1, x_2, \dots, x_n$ независимы, то $M[x_1 \cdot \dots \cdot x_n] = M[x_1] \cdot \dots \cdot M[x_n]$.

Пример 4.

Каждый день местная газета получает заказы на новые рекламные объявления, которые будут напечатаны в завтрашнем номере. Число рекламных объявлений в газете зависит от многих факторов: дня недели, сезона, общего состояния экономики, активности местного бизнеса и т. д.

Найдем математическое ожидание случайной величины X – числа рекламных объявлений в газете в заданный день, если ряд распределения имеет вид:

x_i	0	1	2	3	4	5
$P(x_i) = p_i$	0,1	0,2	0,3	0,2	0,1	0,1

Расчет ожидаемого среднего значения случайной величины удобно производить, пользуясь табл. 2.

Т а б л и ц а 2

Вычисление математического ожидания числа рекламных объявлений

x_i	0	1	2	3	4	5	n
$P(x_i)$	0,1	0,2	0,3	0,2	0,1	0,1	$\sum_{i=1}^n P(x_i)^2$
$x_i P(x_i)$	0,0	0,2	0,6	0,6	0,4	0,5	$M(X) = 2,3$

Можно сказать, что в среднем 2,3 рекламных объявления ежедневно помещаются в газете. Это – **ожидаемое** среднее число рекламных объявлений в заданный день.

В **МП MathCAD** математическое ожидание рассчитывается следующим образом:

$$x := \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \quad p := \begin{pmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.2 \\ 0.1 \\ 0.1 \end{pmatrix}$$

$$M = 2.3$$

В пакете **STADIA** характеристики случайных величин собраны воедино в разделе пакета «Параметрические тесты»/ «Описательная статистика» меню Статистические методы.

- Находясь в электронной таблице пакета, следует ввести данные (рис. 7);
- Выбрать в меню *Статистика*;

Таблица данных		
	x1	x2
1	0	
2	1	
3	2	
4	3	
5	4	
6	5	
7	1	
8	2	
9	2	
10	3	
11		

Рис.7 Ввод данных

- c. В появившемся окне выбрать *Описательная статистика*;
- d. Выбрать переменную таким образом, что бы она переместилась в поле анализа;
- e. Завершить выбор переменных нажав кнопку *Утвердить*.

На экране в окне Результаты появятся значения основных описательных статистик и запрос системы *Выдать дополнительную статистику*. Ответив утвердительно можно получить результаты статистических характеристик.

Результаты							
ОПИСАНИЕ СТАТИСТИКИ. Файл:							
Переменная	Размер	<-- Диапазон-->	Среднее-->	Ошибки	Дисперс	Ст.откл.	Сумма
x1	10	0	5	2,3	0,4726	2,233	23
Переменная	Медиана	<-- Квартили-->	ДоинтСтр.	<ДоинтДисп>	Ош.СтОткл		
x1	2	1 3,25	1,058	1,057	7,445	0,7206	
Переменная	Ассиметр.	Энцисс	Значисс	Значисс			
x1	-0,3032	0,3003	2,3369	0,4546			

Рис.8 Результаты работы программы STADIA

В ТП **Excel** для расчета характеристик можно воспользоваться встроенными функциями из категории «Статистические» или надстройкой программы: Сервис/ Пакет Анализа. Если в меню Сервис отсутствует Пакет Анализа, то выполните следующие действия: Сервис/Надстройки, в появившемся окне поставьте флажок напротив надстройки Пакет Анализа

1. После запуска надстройки необходимо выбрать инструмент анализа *Описательная статистика*;
2. В окне указать входной интервал и ячейку выходного интервала.
3. Отметьте флажком *Итоговая статистика* и нажмите Ок (рис.9);
4. Результаты будут содержать не только значение среднего, но и все характеристики, которые будут рассмотрены ниже.

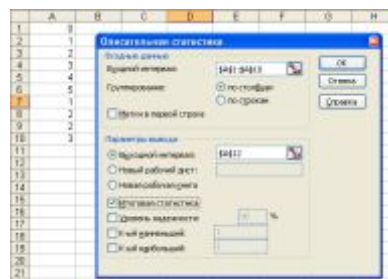


Рис.9 Окно описательной статистики

Расчет математического ожидания для непрерывной случайной величины удобнее выполнять в МП MathCAD.

2. **Дисперсия случайной величины** есть математическое ожидание квадрата отклонения значений случайной величины от ее математического ожидания $D[\mathbf{x}] = M[(\mathbf{x} - M[\mathbf{x}])^2]$.

Справедлива формула $D[\mathbf{x}] = M[\mathbf{x}^2] - M[\mathbf{x}]^2$

Если x – дискретная случайная величина с законом распределения, то

$$D[x] = \int_{-\infty}^{\infty} (x - M[x])^2 f(x) dx, \text{ если } x - \text{непрерывная величина с плотностью } f(x).$$

Пример 5.

Для примера 4(о рекламных объявлениях, размещааемых в газете в определенный день) дисперсия вычисляется так:

$$\sigma^2 = \sum_{i=1}^n [x_i - M(X)]^2 P(x_i) = (0-2,3)^2 + (1-2,3)^2 + (2-2,3)^2 + (3-2,3)^2 + (4-2,3)^2 + (5-2,3)^2 = 2,01.$$

Свойства дисперсии:

1. $D[c] = 0$, если c постоянная.
2. $D[cx] = c^2 D[x] = c$, если c постоянная.
3. Число $\sigma[x] = \sqrt{D[x]}$ называется **средним квадратическое отклонением** x .

Пример 6.

Для примера 4 среднее квадратическое отклонение $\sigma = \sqrt{2,01} = 1,418$.

В чем смысл дисперсии и среднего квадратического отклонения? Как можно интерпретировать их значения? По определению σ^2 – средний квадрат отклонения значений случайной величины от математического ожидания. Отсюда следует, что это **мера рассеяния** всех возможных значений случайной величины относительно среднего ожидаемого значения. Дисперсия характеризует колеблемость, изменчивость случайной величины: чем больше вариация, тем дальше от средней находятся возможные значения случайной величины. Для содержательной интерпретации зачастую полезно применять значение, которое дает корень квадратный из дисперсии – среднее квадратическое отклонение (стандартное отклонение). Если сравнивают две случайные величины, то та из них, которая имеет большую дисперсию и среднее квадратическое отклонение, более вариабельна. Риск, ассоциируемый с инвестициями, часто измеряют стандартным отклонением возврата инвестиций. Если сравниваются два типа инвестиций с одинаковой ожидаемой средней возврата, то инвестиции с более высоким средним квадратическим отклонением считаются более рискованными (хотя более высокое стандартное отклонение предполагает более вариабельный возврат с обеих сторон – как ниже, так и выше средней).

Задание 3 Дискретная случайная величина x задана законом распределения:

Значения x	-1	0	1	2
Вероятности	0,1	0,3	0,5	0,1

Найти математическое ожидание величин x и $y = 2x^2 + 1$ и дисперсию.

$$M[x] = \underline{\hspace{10cm}}$$

$$M[y] = M[2x^2 + 1] = \underline{\hspace{10cm}}$$

$$D[x] = \underline{\hspace{10cm}}$$

$$D[y] = D[2x^2 + 1] = \underline{\hspace{10cm}}$$

Задание 4 Случайная величина x имеет плотность вероятности

$$f(x) = \begin{cases} 0, & \text{если } x < 0, \\ \alpha e^{-\alpha x}, & \text{если } x \geq 0, \alpha > 0. \end{cases}$$

Найти математические ожидания, дисперсию и среднее квадратическое отклонение величин x .

$$M[x] = \int_0^{\infty} x \alpha e^{-\alpha x} dx = \underline{\hspace{10cm}}$$

$$D[x] = \int_0^{\infty} x^2 \alpha e^{-\alpha x} dx = \underline{\hspace{10cm}}$$

4. **Начальным моментом k -го порядка** (m_k) случайной величины X называется математическое ожидание ее k -й степени:

$$\text{для дискретной случайной величины } m_k = \sum_{i=1}^n x_i^k p_i;$$

для непрерывной случайной величины $m_k = \int_{-\infty}^{\infty} x^k W(x) dx$.

5. **Центральным моментом k -го порядка** (μ_k) случайной величины X называют математическое ожидание k -й степени отклонения случайной величины X от ее математического ожидания:

для дискретной случайной величины $\mu_k = \sum_{i=1}^n [x_i - M(X)]^k p_i$;

для непрерывной случайной величины $\mu_k = \int_{-\infty}^{\infty} [x - M(X)]^k W(x) dx$

Заметим, что начальный момент первого порядка m_1 представляет собой математическое ожидание случайной величины, а центральный момент второго порядка μ_2 – дисперсию случайной величины.

6. Центральный момент третьего порядка применяется для характеристики скошенности или асимметрии распределения (**коэффициент асимметрии**): $\beta_1 = \mu_3/\sigma^3$

Для симметричных распределений $\beta_1 = 0$.

7. Центральный момент 4-го порядка применяется для характеристики крутости (эксцесса) распределения (**неприведенный коэффициент эксцесса**): $\beta_2 = \mu_4/\sigma^4$.

Пример 7 Для примера 4 определить характеристики случайной величины X рассмотренные в пунктах 2-7.

- в **ТП Excel** возможно с помощью Пакета Анализа (рис.11), а так же с помощью встроенных функций:

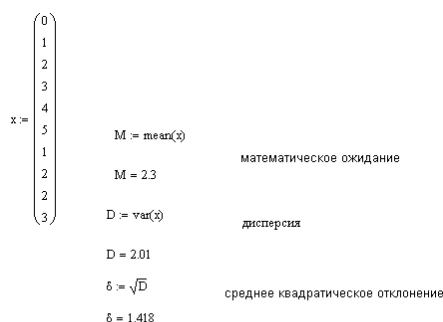
СКОС(число1;число2; ...) - возвращает асимметрию распределения. Асимметрия характеризует степень несимметричности распределения относительно его среднего.

ЭКСЦЕСС(число1;число2; ...) - эксцесс характеризует относительную остроконечность или сглаженность распределения по сравнению с нормальным распределением.

ДИСПР(число1;число2; ...) - дисперсия

СТАНДОТКЛОНП(число1; число2; ...) - средне квадратичное отклонение;

- В статистическом пакете **STADIA** (рис 8);
- В МП **MathCAD** имеются встроенные функции среднего значения - $\text{mean}(x)$; дисперсии - $\text{var}(x)$.



Или с помощью формул:

$$x := \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \quad p := \begin{pmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.2 \\ 0.1 \\ 0.1 \end{pmatrix}$$

+

$$M := \sum (x \cdot p) \quad M = 2.3 \quad \text{математическое ожидание}$$

$$M2 := \sum (x^2 \cdot p)$$

$$D := M2 - M^2 \quad D = 2.01 \quad \text{дисперсия}$$

$$\mu_1 := \sum (x \cdot p) \quad \mu_1 = 2.3 \quad \text{начальный момент первого порядка}$$

$$\mu_2 := \sum (x^2 \cdot p) \quad \mu_2 = 7.3 \quad \text{начальный момент второго порядка}$$

$$\lambda_1 := \sum [(x - M)^1 \cdot p] \quad \lambda_1 = 0 \quad \text{центральный момент первого порядка}$$

$$\lambda_2 := \sum [(x - M)^2 \cdot p] \quad \lambda_2 = 2.01 \quad \text{центральный момент второго порядка}$$

$$\lambda_3 := \sum [(x - M)^3 \cdot p] \quad \lambda_3 = 0.864 \quad \text{центральный момент третьего порядка}$$

+

ПРАКТИЧЕСКАЯ РАБОТА №2

Расчет числовых характеристик случайных величин

Цель работы: научиться решать задачи на нахождение числовых характеристик случайной величины.

Порядок выполнения работы:

1. Задачу 1 решить в статистическом пакете STADIA(по вариантам);
2. Задачу 2 решить в МП MathCAD;
3. Задачу 3 решить в МП MathCAD (по вариантам);
4. В ТП
- 5.

Задача 1. Дискретная случайная величина x задана таблицей распределения. Требуется найти $M[x]$, $D[x]$ и $\sigma[x]$:

1)	0	1	2	3)	-2	-1	0	1	2
	0,3	0,5	0,2		0,1	0,2	0,4	0,2	0,1

2)	-1	0	1	4)	1	2	3	4	5
	0,4	0,5	0,1		0,05	0,15	0,3	0,4	0,1

5)	-2	0	2	6)	-2	-1	0	1	2
	0,3	0,4	0,3		0,05	0,15	0,3	0,4	0,1

7)	1	2	3	8)	0	1	2	3	4
	0,4	0,5	0,1		0,1	0,1	0,3	0,2	0,3

9)	2	5	8	10)	2	4	6	8	10
	0,3	0,4	0,3		0,7	0,05	0,1	0,05	0,1

Задача 2.

Строительная инвестиционная компания в настоящий момент продает акции по 16 условных денежных единиц за штуку. Инвестор планирует покупку пакета акций и предполагает хранение их в течение года. Пусть X – случайная величина, означающая цену одной акции спустя год. Ряд распределения дан в таблице:

Цена акции (x)	$P(X)$
16	0,35
17	0,25
18	0,25
19	0,10
20	0,05

1. Показать, что заданное распределение обладает всеми свойствами ряда распределения.
2. Чему равно ожидаемое среднее значение цены акции спустя один год?
3. Чему равен ожидаемый средний выигрыш от акции, спустя год? Чему равен процент возврата инвестиций, отражаемый этим ожидаемым значением?
4. Определите дисперсию цены акции спустя год.
5. Другая акция с одинаковым ожидаемым значением возврата инвестиций имеет дисперсию, равную 3. Какая из акций лучше в смысле минимизации риска или неопределенности, ассоциируемой с инвестициями? Объясните.

Задача 3. Непрерывная случайная величина x задана плотностью вероятности $f(x)$. Найдите $M[x]$, $D[x]$ и $\sigma[x]$.

$$1) f(x) = \begin{cases} 0, & \text{если } x < 0, \\ 2x, & \text{если } 0 \leq x \leq 1, \\ 0, & \text{если } x > 1; \end{cases} \quad 2) f(x) = \begin{cases} 0, & \text{если } x < 2, \\ 0,5x, & \text{если } 2 \leq x \leq 4, \\ 0, & \text{если } x > 4; \end{cases}$$

$$3) f(x) = \begin{cases} 0, & \text{если } x < -\frac{\pi}{2}, \\ 0,5\cos, & \text{если } -\frac{\pi}{2} \leq x \leq \frac{\pi}{4}, \\ 0, & \text{если } x > \frac{\pi}{2}; \end{cases} \quad 4) f(x) = \begin{cases} e^x, & \text{если } x \leq 0, \\ 0, & \text{если } x > 0. \end{cases}$$

$$5) f(x) = \begin{cases} 0, & \text{если } x < 0, \\ x/2, & \text{если } 0 \leq x \leq 1, \\ 0, & \text{если } x > 1; \end{cases} \quad 6) f(x) = \begin{cases} 0, & \text{если } x < 2, \\ 0,5x^2, & \text{если } 2 \leq x \leq 4, \\ 0, & \text{если } x > 4; \end{cases}$$

$$7) f(x) = \begin{cases} 0, & \text{если } x < -\frac{\pi}{4}, \\ 3\cos, & \text{если } -\frac{\pi}{4} \leq x \leq \frac{\pi}{4}, \\ 0, & \text{если } x > \frac{\pi}{2}; \end{cases} \quad 8) f(x) = \begin{cases} \cos(x), & \text{если } x > 0, \\ 0, & \text{если } x \leq 0. \end{cases}$$

$$9) f(x) = \begin{cases} 0, & \text{если } x < 0, \\ x/5, & \text{если } 0 \leq x \leq 5, \\ 0, & \text{если } x > 5; \end{cases} \quad 10) f(x) = \begin{cases} 0, & \text{если } x < 2, \\ 5x, & \text{если } 2 \leq x \leq 5, \\ 0, & \text{если } x > 5; \end{cases}$$

Задача 4.

Некоторое предприятие планирует реконструкцию и расширение производства для выпуска новой продукции. Руководство предприятия должно определить стратегию реконструкции и выбрать один из двух проектов, предусматривающих большие и умеренные капитальные вложения. Неопределенность заключается в том, что спрос на новую продукцию, которую собирается выпускать предприятие, неизвестен. Будущий спрос может быть низким, умеренным и высоким. Вероятности спроса оцениваются как 0,20, 0,50 и 0,30 соответственно. Пусть X означает ежегодный доход 1000 условных денежных единиц. Предприятие планирует следующий доход для проектов с большими и умеренными капитальными вложениями:

Доход при значительных вложениях		Доход при умеренных вложениях	
Спрос, x	$P(x)$	Спрос, x	$P(x)$
0	0,20	50	0,20
100	0,50	150	0,50
300	0,30	200	0,30

1. Вычислите ожидаемое среднее значение дохода при альтернативных типах реконструкции предприятия. Какое решение предпочтительнее для максимизации ожидаемого дохода?

2. Вычислите дисперсию дохода для двух альтернативных проектов. Какое решение Вы предпочтете для минимизации риска и неопределенности?

Тест для самоконтроля.

1. Выберите более полное и правильное определение случайной величины:

- а) случайной называется величина, которая в результате опыта может принять то или иное возможное значение, неизвестное заранее, но обязательно одно;
- б) случайной называется величина, которая в результате опыта может принять одно заранее определенное значение, но обязательно одно;
- в) случайной называется величина, которая в результате опыта может принять всевозможные значения из множества допустимых значений.

2. Запишите функцию распределения для следующей задачи:

Два игрока подбрасывают монету. Если при одном подбрасывании выпал «герб», то первый игрок получает 10 рублей, а если выпала «решка» то отдает.

3. Продолжите предложение. Функцией распределения случайной величины называется ...

4. Запишите формулу связи функции и плотности распределения.

5. Задан закон распределения:

X	6	3	1
P	0,2	0,3	0,5

математическое ожидание дискретной случайной величины равно:

- а) 6
- б) 7,3
- в) 2,6

6. Докажите свойство $M(x+y) = M(x) + M(y)$

7. Запишите формулу для расчета дисперсии.

8. Какому значению равен центральный момент второго порядка?

Тема №5 Виды распределений дискретной и непрерывной случайной величины

Изучив эту тему, Вы будете:

Иметь представление:

- о видах случайных величин, о законах распределения случайной величины;

Знать:

- виды распределений случайных величин.

Уметь:

- применять различные виды распределений для нахождения вероятности случайных величин, композиции законов распределения.

Входной тест

1. Продолжите предложение. Распределением дискретной случайной величины называется ...
2. Чему равно значение функции распределения в точке $+\infty$
3. Выберите из предложенных формул MathCAD, формулу для расчета числа комбинаций из n различных элементов по k :
 - а). $\text{combin}(n,k)$
 - б). $\text{permut}(n,k)$
 - в). $\text{combin}(k,n)$
 - г). $\text{permut}(k,n)$

5.1 Биномиальное распределение

Испытания Бернулли – это последовательность n идентичных испытаний, удовлетворяющих следующим условиям:

1. Каждое испытание имеет два исхода: успех и неуспех – взаимно несовместные и противоположные события.
2. Вероятность успеха p остается постоянной от испытания к испытанию. Вероятность неуспеха $q = 1-p$.
3. Все n испытаний – независимы. Вероятность наступления события в любом из испытаний не зависит от результатов других испытаний.

Успех и неуспех – статистические термины. Например, когда имеют дело с производственным процессом, то исход испытания «деталь дефектная» определяют как успех. Успех относится к появлению определенного события – «деталь дефектная», а неуспех относится к неоявлению события. Определим случайную величину как биномиальную, если для нее мы рассчитываем число успехов и неуспехов в последовательности n испытаний Бернулли.

Если производить n независимых опытов, в каждом из которых событие A (которое условно можно назвать «успехом» опыта) появляется с вероятностью p , то случайная величина X – число «успехов» при n опытах имеет биномиальное распределение.

В этом случае событие $A=\{X=m\}$ распадается на ряд вариантов, в каждом из которых «успех» достигается в m опытах, а «неуспех» в $(n-m)$ опытах.

По правилу умножения вероятностей $P(A) = p^m (1-p)^{n-m}$ или обозначая $q=1-p$, $P(A) = p^m q^{n-m}$.

Таким образом, дискретная случайная величина имеет биномиальное распределение, если ее возможные значения имеют величину $0, 1, \dots, m, \dots, n$, а соответствующие вероятности принимают значения $P\{X = m\} = C_n^m p^m q^{n-m}$, где $0 < p < 1$; $q=1-p$; $m=0, 1, \dots, n$.

Такое распределение зависит от двух параметров n и p .

Пример №1 Монета подбрасывается 4 раза, пусть X – число появившихся гербов.

$n = 4$, $p = 0,5$ – параметры биномиального распределения случайной величины X . Последовательные подбрасывания монеты – независимые эксперименты; исходы – «цифра» или «герб» (успех – неуспех) и вероятности их выпадения постоянны от испытания к испытанию.

Пример №2 Известно, что в определенном городе 30 % горожан предпочитают добираться на работу личным автотранспортом. Случайно выбраны 8 человек. Пусть Y – число людей в выборке, предпочитающих личный автотранспорт.

$n = 8, p = 0,3$ – параметры биномиального распределения случайной величины Y . Заметим, что случайная выборка из большой генеральной совокупности предполагает независимость испытаний. Мы полагаем, что число людей в городе (генеральная совокупность) намного больше, чем число испытаний, и случайный отбор небольшого числа людей не влияет на ту часть оставшихся горожан, которые предпочитают добираться до работы на личном транспорте (события «предпочитают личный транспорт» для любых отобранных горожан – независимы). Если в генеральной совокупности только 10 человек, трое из которых предпочитают личный транспорт, то ситуация меняется. Вероятность того, что следующий отобранный горожанин предпочтет также личный транспорт, составит уже только $2/9 \approx 0,22$ или $3/9 \approx 0,33$ в зависимости от того, предпочитает ли он личный транспорт или нет. В этом случае условия 2 и 3 испытаний Бернулли будут нарушены и Y не будет биномиальной случайной величиной. Чем больше объем генеральной совокупности в сравнении с выборкой, тем менее серьезно нарушение условий 2 и 3. На практике пользуются правилом: если $N/n > 10$ (N – объем генеральной совокупности, n – объем выборки), то можно предположить независимость исходов.

Задание №1

Известно, что 15 % деталей, произведенных автоматом, – бракованные. В порядке случайного отбора взято 12 деталей.

Определить параметры биномиального распределения числа появления деталей без брака: $n =$ _____, $p =$ _____

Определение. Дискретная случайная величина X , которая может принимать только целые неотрицательные значения с вероятностью

$$P_n(m) = P(X = m) = C_n^m p^m q^{n-m},$$

где $p + q = 1, p > 0, q > 0, m = 0, 1, 2, \dots, n$, называется *распределенной по биномиальному закону*, а p – параметром биномиального распределения.

Ряд распределения случайной величины, подчиненной биномиальному закону, можно представить в следующем виде:

$X = m$	0	1	...	k	...	n
$P_n(m)$	$C_n^0 p^0 q^n$	$C_n^1 p^1 q^{n-1}$	...	$C_n^k p^k q^{n-k}$	...	$C_n^n p^n q^0$

Функция распределения в этом случае определяется формулой

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ \sum_{m \leq x} C_n^m p^m q^{n-m} & \text{при } 0 < x \leq n, \\ 1 & \text{при } x > n. \end{cases}$$

Числовые характеристики этого распределения:

- 1) $MX = np$.
- 2) $DX = npq$, или $Dm = npq$.
- 3) $\sigma_x = \sqrt{npq}$, или $\sigma_m = \sqrt{npq}$.
- 4) $A = \frac{q-p}{\sqrt{npq}}$.
- 5) $E = \frac{1-6pq}{npq}$.

Биномиальное распределение асимметрично, за исключением случая, когда $p = 0,5$, причем при $p < 0,5$ асимметрия положительна, а при $p > 0,5$ – отрицательна.

Пример №3

Для примера 1 биномиальное распределение X – числа гербов, появляющихся при четырех подбрасываниях монеты равно:

Таблица 3

$X = m$	0	1	2	3	4
$P(x) = P_{4,m}$	0,0625	0,2500	0,375	0,2500	0,0625

С увеличением числа испытаний расчет вероятностей по формуле становится все более громоздким. Существуют специальные таблицы, в которых табулированы значения вероятностей биномиального распределения для различных n и p . Иногда в литературе предлагаются таблицы, в которых табулированы значения интегральной функции $1-F(x) = P(X \geq x)$. Табл. 4 воспроизводит значения функции при $n = 4$. Найдем кумулятивную вероятность, которой соответствует распределение, представленное в табл. 3. Заметим, что для $p = 0,5$

$$P(X = 2) = \sum_{i=0}^2 P(i) - \sum_{i=0}^1 P(i) = F(2) - F(1) = 0,6875 - 0,3125 = 0,3750,$$

т. е. в общем виде: $P(X) = F(x) - F(x-1)$

Вероятность, равная 0,3750, корреспондирует с вероятностью при $m = 2$ в табл. 4.

Таблица 4

Фрагмент таблицы $F(x) = P(X \leq x)$ биномиального распределения

m	0	1	2	3	4
$P(X \leq x) = F(x)$	0,06250	0,3125 = 0,0625 + + 0,2500	0,6875 = 0,3125 + + 0,3750	0,9375 = 0,6875 + + 0,2500	1,0000 = 0,9375 + + 0,0625

Для решения подобных задач можно воспользоваться ТП Excel или МП MathCAD.

МП MathCAD	ТП Excel
$P(X=k) =$ $g(k,n,p) =$ $dbinom(k,n,p)$ $:= \left(\frac{n}{k}\right) p^k (1-p)^{n-k}$ P — это вероятность того, что в выборке объема n будет находиться ровно k дефектных элементов, если доля дефектных элементов в генеральной совокупности составляет p (предполагается, что n намного меньше, чем число элементов в генеральной совокупности).	БИНОМРАСП(число_успехов;число_испытаний;вероятность_успеха;интегральная) Число_успехов — это количество успешных испытаний. Число_испытаний — это число независимых испытаний. Вероятность_успеха — это вероятность успеха каждого испытания. Интегральная — это логическое значение, определяющее форму функции. Если аргумент интегральная имеет значение ИСТИНА, то функция БИНОМРАСП возвращает интегральную функцию распределения; если этот аргумент имеет значение ЛОЖЬ, то возвращается функция распределения.

Пример №4

Баскетболист выполнил серию из 3 штрафных бросков, причем вероятность попадания в кольцо для каждой из попыток равна 0,5. Определить вероятность того, что в данной серии баскетболист поразит кольцо ровно: 0,1,2,3 раза.

Решение выполнено в МП MathCAD:

$n := 3$ количество бросков
 $p := 0.5$ вероятность попадания при одном броске
 $x := 0..n$ варианты попадания баскетболистом в кольцо
 $p_x := \frac{n!}{x!(n-x)!} \cdot p^x \cdot (1-p)^{n-x}$
 биномиальный закон распределения величины x
 $p = \begin{bmatrix} 0.125 \\ 0.375 \\ 0.375 \\ 0.125 \end{bmatrix}$
 проверка $\sum_x p_x = 1$

Задание №2

Для данных из примера 2 найдите вероятности того, что предпочтут личный транспорт: а) 5 человек из 8; б) не более 5 человек; в) не менее 5 человек. По условию $p = 0,3$. Значит, надо определить $P(X = 5)$, $P(X \leq 5)$, $P(X \geq 5)$. Результаты занесите в таблицу 5.

Таблица 5

Таблица ряда и функции биномиального распределения.

$X = m$	$P(X = m) = C_n^m p^m q^{n-m}$	$P(X \leq m) = F_1(x)$	$P(X < x) = F(x)$	$P(X \geq x) = 1 - F(x)$

5.2 Распределение Пуассона

Распределение Пуассона (закон распределения редких событий) часто используется тогда, когда мы имеем дело с числом событий, появляющихся в промежутке времени или пространства (число машин, прибывших на автомойку в течение часа, число дефектов на новом отрезке шоссе длиной в 10 км, число мест утечки воды на 100 км водопровода, число остановок станков в неделю, число дорожных происшествий).

Если вероятность появления события A в n отдельных независимых испытаниях очень мала ($p < q$), то применяется формула Пуассона:

$$P_{m,n} \approx \frac{\lambda^m}{m!} \cdot e^{-\lambda},$$

где $\lambda = np$; n – число независимых испытаний с постоянной малой вероятностью p ; e – основание натурального логарифма ($e = 2,71828$);

m – число появлений события ($m = 0, 1, 2, 3, \dots$).

Закон распределения Пуассона можно написать в виде ряда распределения (табл. 6), если, придавая m целые неотрицательные значения $m = 0, 1, 2, \dots, n$, вычислить соответствующие им вероятности $P_{n,m}$.

Таблица 6

m	0	1	2	3	...	k	...	n
$P_{n,m}$	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\lambda^2 e^{-\lambda}/2!$	$\lambda^3 e^{-\lambda}/3!$	...	$\lambda^k e^{-\lambda}/k!$	...	$\lambda^n e^{-\lambda}/n!$

Закон распределения Пуассона можно записать в виде функции распределения: $\lambda^k e^{-\lambda}/k!$

$$F(X) = P(m < x) = \sum_{m < x} P_{n,m} = \sum_{m < x} \lambda^m / k! e^{-\lambda},$$

$$F(X) = \sum_{m < x} \frac{\lambda^m}{m!} e^{-\lambda}, \text{ где знак } \sum_{m < x} \text{ означает сумму вероятностей } P_{n,m} \text{ для всех } m, \text{ меньших } n.$$

Определить вероятность появления события хотя бы один раз в n независимых испытаниях возможно применяя формулу:

$$P_{n,m \geq 1} = 1 - P_{n,0} = 1 - \frac{\lambda^0}{0!} \cdot e^{-\lambda} = 1 - e^{-\lambda}.$$

Математическое ожидание и дисперсия случайной величины, распределенной по закону Пуассона, совпадают и равны параметру λ , который определяет этот закон, т. е. $M(X) = D(X) = \lambda$.

Эта Формула устанавливает важный теоретико-вероятностный смысл параметра λ . Последовательность событий, которые наступают в случайные моменты времени, называется **потоком событий** (например, вызов на АТС).

При этом должны выполняться следующие условия.

Вероятность появления события одна и та же для любых двух интервалов равной длины.

Вероятность того, что событие появится в короткий интервал времени (или пространства), пропорциональна величине интервала.

В очень коротком интервале вероятность того, что два события появятся, близка к нулю.

Вероятность того, что любое число событий появится в интервале, не зависит от начала интервала.

Появление или не появления события в определенном интервале не зависит от появления или не появления события в любом другом интервале.

Пример №5 Предположим, нас интересует число инкассаторов, прибывающих утром на автомобиле в банк в течение 15 мин. Если мы предположим, что вероятность прибытия автомобиля одинакова в любые два периода времени равной длины и что прибытие или неприбытие автомобиля в любой период времени не зависит от прибытия или неприбытия в любой другой период времени, то последовательность прибытия инкассаторов в банк может быть описана распределением Пуассона.

Анализ предыдущих данных показал, что среднее число инкассаторов, прибывающих в 15-минутный период, равно 10, тогда при $\lambda = 10$ получаем: $P(m) = \lambda^m e^{-\lambda} / m! = 10^m e^{-10} / m!$ при $m = 0, 1, 2, \dots$

Если мы хотим узнать вероятность прибытия пяти инкассаторов в течение 15 мин, то при $m = 5$ получим: $P(5) = 10^5 e^{-10} / 5! = 0,0378$.

Вероятности распределения Пуассона легче рассчитать, пользуясь специальными таблицами вероятностей распределения Пуассона или ТП Excel, МП Matcad.

МП MathCAD			ТП Excel
$P(X=k)$	=	$g(k, \mu)$	ПУАССОН(х;среднее;интегральная)
$dpois(k, \mu) := \frac{\mu^k}{k!} \cdot e^{-\mu}$			х — количество событий. Среднее — ожидаемое численное значение. Интегральная — логическое значение, определяющее форму возвращаемого распределения вероятностей. Если аргумент «интегральная» имеет значение ИСТИНА, то функция ПУАССОН возвращает интегральное распределение Пуассона. Если этот аргумент имеет значение ЛОЖЬ, то возвращается функция плотности распределения Пуассона.
P — это вероятность того, что при испытании устройства будет зафиксировано ровно k отказов, когда среднее число отказов на одно устройство в генеральной совокупности имеет значение μ .			

Задание №3

Проведено 25 независимых испытаний с вероятностью появления события A в каждом из них 0,01. Постройте ряд распределения для случайной величины $X = m$ — числа появлений события A . Вероятность $P_{n,m}$ вычислите двумя способами: по формуле Бернулли и по формуле Пуассона. Полученные результаты занесите в таблицу №7. Сравните и оцените погрешности приближенной формулы. По условию $n = 25$; $p = 0,01$; $q = 0,99$.

Таблица 7

Сравнение вероятностей, полученных по формулам Бернулли и Пуассона.

m	0	1	2	3	4	5	6
$P_{n,m} = C_n^m p^m q^{n-m}$							
$P_{n,m} = \frac{\lambda^m}{m!} e^{-\lambda}$							
$ \Delta $							

Сопоставление вероятностей показывает, что рассчитанные по формуле Пуассона и Бернулли вероятности _____.

Максимальная погрешность результатов, вычисленных по формуле Пуассона, равна _____.

5.3 Нормальное распределение

Наиболее важным распределением непрерывных случайных величин является нормальное распределение. Множество явлений в практической жизни можно описать с помощью модели нормального распределения, например распределение высоты деревьев, площадей садовых участков, массы людей, дневной температуры и т. д. Нормальное распределение используется и для решения многих проблем в экономической жизни. Например, распределение числа дневных продаж в магазине, числа посетителей универсама в неделю, числа работников в некоторой отрасли, объемов выпуска продукции на предприятии и т. д.

Нормальное распределение иногда называют **законом ошибок**. Например, отклонения в размерах деталей от установленного объясняются многими причинами, каждая из которых влияет на размер детали, так что отклонение, которое фактически регистрируется при измерениях, является суммой большого числа отклонений (ошибок).

Нормальная случайная величина имеет плотность распределения, определяемую формулой

$$W(x) = \frac{1}{\lambda\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\lambda^2}}, \text{ где } -\infty < x < +\infty, a = M(X), \lambda = \sigma(X).$$

Основные свойства $W(x)$:

а) функция $W(x)$ существует при любых действительных значениях x и принимает только положительные значения. Следовательно, нормальная кривая распределения расположена выше оси абсцисс;

б) при неограниченном возрастании x по абсолютной величине $W(x)$ стремится к нулю, значит, ось абсцисс служит горизонтальной асимптотой кривой нормального распределения;

в) максимальное значение функция $W(x)$ принимает в точке, соответствующей математическому ожиданию случайной величины x .

Действительно, приравнявая первую производную от $W(x)$ к нулю, т. е.

$$W'(x) = -\frac{x-a}{\lambda^3\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\lambda^2}} = 0,$$

убеждаемся в том, что $W'(x) = 0$ при $x = a$; $W'(x) > 0$ при $x < a$; $W'(x) < 0$ при $x > a$.

Следовательно, функция $W(x)$ принимает максимальное значение в точке $x = a$;

г) кривая нормального распределения симметрична относительно прямой $x = a$

д) кривая нормального распределения имеет две точки перегиба, симметрично расположенные относительно прямой $x = a$, с абсциссами точек перегиба $a - \lambda$ и $a + \lambda$ и ординатами $1/(\lambda\sqrt{2\pi})$.

Функция нормального распределения содержит два параметра: математическое ожидание $a = M(X)$ и стандартное отклонение $\lambda = \sigma(X)$. Следовательно, существует бесконечно много нормально распределенных случайных величин, у которых разные $M(X)$ и $\sigma(X)$. Графики их плотностей имеют одинаковую форму – симметричную, колоколообразную. Если $M(X)$ и $\sigma(X)$ известны, то из семейства нормальных случайных величин выделяем конкретную нормальную случайную величину с определенной плотностью (таблица 8).

Таблица 8

Математическое ожидание a и стандартное отклонение λ	
1. Это величина, которая характеризует положение кривой распределения на оси абсцисс.	
2. Изменение параметра a при неизменном λ приводит к перемещению оси симметрии ($x = a$) вдоль оси абсцисс и, следовательно, к соответствующему перемещению кривой распределения.	
Изменение среднего квадратического отклонения при фиксированном значении	

М(х) приводит к изменению формы кривой распределения.	
---	--

Аналогично, как и для распределений дискретной случайной величины расчет функции и плотности распределения для нормальной случайной величины можно осуществлять в ТП Excel или МП MathCAD.

МП MathCAD	ТП Excel
$g(x, \mu, \sigma)$ $\text{dnorm}(x, \mu, \sigma) := \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{\frac{(x-\mu)^2}{2 \cdot \sigma^2}}$ а функ- цию распре- деления $G(x) := \int_{-\infty}^x g(t) dt$ рассчитать по формуле:	ПУАССОН(х;среднее;интегральная) х — количество событий. Среднее — ожидаемое численное значение. Интегральная — логическое значение, определяющее форму возвращаемого распределения вероятностей. Если аргумент «интегральная» имеет значение ИСТИНА, то функция ПУАССОН возвращает интегральное распределение Пуассона. Если этот аргумент имеет значение ЛОЖЬ, то возвращается функция плотности распределения Пуассона.

Вероятность попадания в заданный интервал нормально распределенной случайной величины

Если случайная величина задана плотностью распределения $W(x)$, то вероятность того, что X примет значение, принадлежащее интервалу (α, β) , определяется из выражения

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} W(x) dx$$

Если случайная величина $X \sim N(a; \sigma^2)$, то

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

Для расчета вероятности попадания в заданный промежуток удобнее воспользоваться **интегралом вероятностей**, или **функцией Лапласа**. Его обычно обозначают символом $\Phi_0(z)$:

$$\Phi_0(z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Интеграл Лапласа в общем виде не берется. Его можно вычислить одним из способов приближенного вычисления интегралов. Эта функция табулирована. Пользуясь функцией Лапласа, окончательно получаем:

$$P(\alpha < X < \beta) = \Phi_0\left(\frac{\beta-a}{\lambda}\right) - \Phi_0\left(\frac{\alpha-a}{\lambda}\right).$$

Пример № 6

Найдем вероятность того, что нормально распределенная случайная величина Z примет значение, принадлежащее интервалу от 1 до 2.

Искомая вероятность – это площадь под кривой между точками 1 и 2 (рис.12) $P(1 < Z < 2) = \Phi_0(2) - \Phi_0(1) = 0,47725 - 0,34134 = 0,13591$.

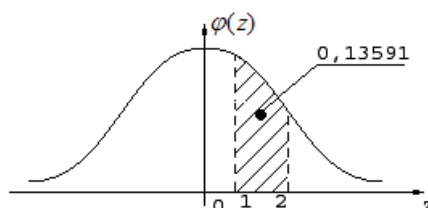


Рисунок 12 Вероятность нормально распределенной случайной величины.

Задание № 4

Найдите и запишите вероятность того, что стандартная нормально распределенная случайная величина будет иметь значения между -2 и 1.

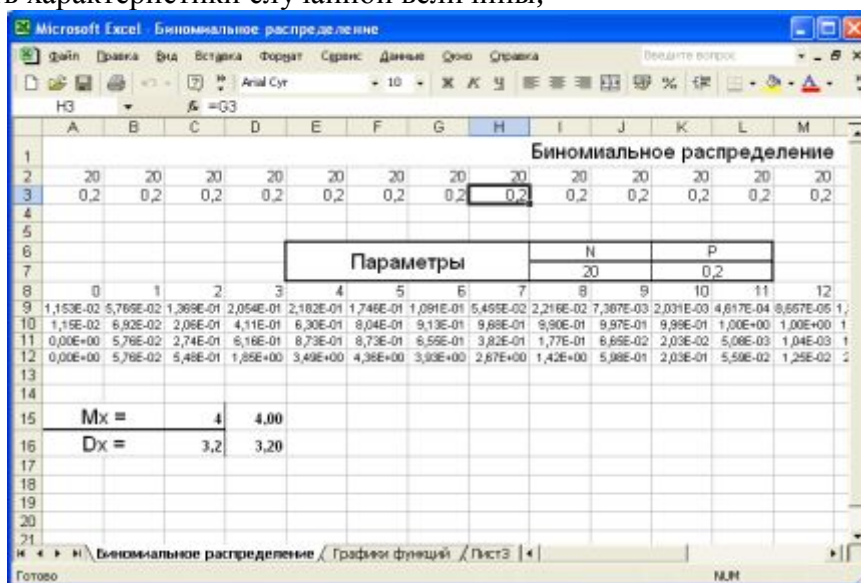
$$P(-2 < Z < 1) = \Phi_0(-2) - \Phi_0(1) = \underline{\hspace{2cm}}$$

Практическая работа № 3 Законы распределения вероятностей в пакетах STADIA, Excel, MathCAD

1. Биномиальное распределение (ТП Excel).

Порядок выполнения работы:

- Составить ряд распределения дискретной случайной величины согласно индивидуальному заданию практической работы;
- Построить многоугольник распределения и график функции распределения дискретной случайной величины;
- Вычислить характеристики случайной величины;



В первой строке таблицы введено число, показывающее количество опытов, которые необходимо провести (согласно индивидуального задания). Во второй строке – соответствующие значения вероятностей.

В графе «Параметры» эти значения введены как глобальные параметры для всех опытов. При этом N – характеризует количество опытов, а значение p – вероятность этих опытов.

В восьмой строке числа соответствуют возможным значениям случайной величины. Порядок вычисления функции биномиального распределения в среде Excel следующий:

БИНОМРАСП(число_успехов;число_испытаний;вероятность_успеха ;интегральная)

Число_успехов — это количество успешных испытаний.

Число_испытаний — это число независимых испытаний.

Вероятность_успеха — это вероятность успеха каждого испытания.

Пример:

	А	В
	Данные	Описание
1.	6	Количество успешных испытаний
2.	10	Число независимых испытаний
3.	0,5	Вероятность успеха в каждом испытании
4.	=БИНОМРАСП(А2;А3;А4;ЛОЖЬ)	Вероятность того, что в точности 6 испытаний из 10 будут успешны 0,205078

По полученным данным построены графики функций:

Значения числовых характеристик случайной величины, такие как математическое ожидание и дисперсия рассчитать с использованием встроенных функций. В приведенном примере результаты этих двух расчетов приведены в пятнадцатой и шестнадцатой строках таблицы. На листе 3 представлены значения характеристик для сгенерированной случайной величины с помощью пакета Анализа. Этапы проведения показаны на рисунке 13.

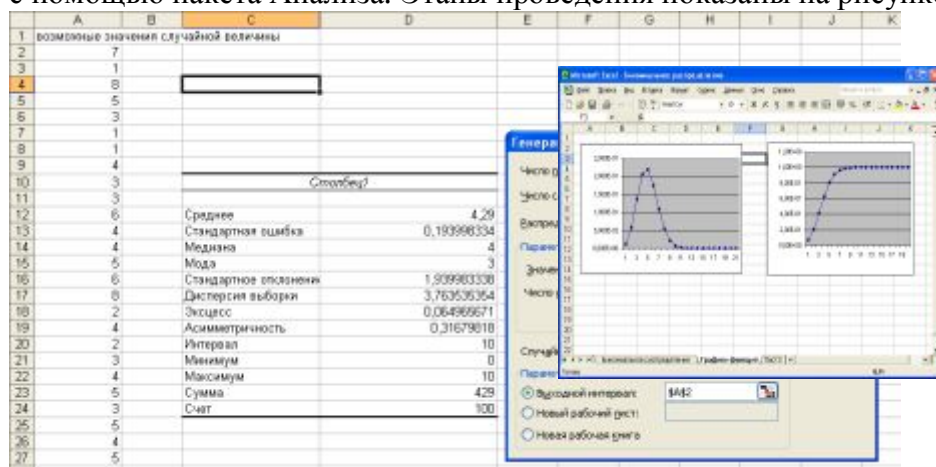


Рисунок 13 Генерация случайных чисел

Индивидуальное задание:

Таблица 9

№ компьютера	Число испытаний	Вероятность
1	30	0,2
2	50	0,5
3	70	0,7
4	90	0,9
5	110	0,5
6	90	0,2
7	70	0,1
8	50	0,6
9	30	0,7
10	20	0,3

2. Распределение Пуассона (MPI MathCAD).

Порядок выполнения работы:

- Введите готовое решение представленной ниже задачи;
- Повторите все вычисления для данных из индивидуального задания, изменив в пунктах 1) и 3) соответствующие значения;
- Сравните полученные результаты.

Провайдер обслуживает 1000 абонентов сети INTERNET. Вероятность того, что любой абонент захочет войти в сеть в течение часа, равна 0.002. Найти вероятность того, что в течение часа более 10 абонентов попытаются войти в сеть.

Решение:

1) Введем число n обслуживаемых абонентов: $n := 1000$

2) Введем число k_0 обслуживаемых абонентов в течение часа: $k_0 := 10$

3) Введем вероятность p того, что любой абонент захочет войти в сеть в течение часа: $p := 0.002$

4) Определим вероятность того, что в течение часа ровно k_0 абонентов попытается войти в сеть (формула Бернулли):

$$C(1, n) := \frac{n!}{1! \cdot (n - 1)!}$$

$$Pb1(k) := C(k, n) \cdot p^k \cdot (1 - p)^{n-k}$$

$$Pb1(k0) \rightarrow 3.71678435838646277210^{-5}$$

5) Определим вероятность того, что в течение часа менее k0 абонентов попытаются войти в сеть (функция распределения):

$$Pb2(k) := \sum_{i=0}^{k-1} Pb1(i)$$

$$Pb2(k0) \rightarrow .99995482531168377:$$

6) Определим требуемую вероятность:

$$Pb3(k) := 1 - Pb1(k) - Pb2(k)$$

$$Pb3(k0) \rightarrow 8.0068447323567100^{-6}$$

7) Определим требуемую вероятность, используя приближенную формулу Пуассона:

$$Pp1(k) := \frac{(n \cdot p)^k}{k!} \cdot e^{-n \cdot p}$$

$$Pp1(k0) \rightarrow 3.81898506487795955910^{-5}$$

$$Pp2(k) := \sum_{i=0}^{k-1} Pp1(i)$$

$$Pp2(k0) \rightarrow .99995350192498273:$$

$$Pp3(k) := 1 - Pp1(k) - Pp2(k)$$

$$Pp3(k0) \rightarrow 8.3082243684842410^{-6}$$

3. Нормальное распределение (МП MathCAD).

Порядок выполнения работы:

- Постройте графики плотности нормального распределения и функции нормального распределения $N(0,1)$ (стандартное нормальное распределение) значение параметров распределения возьмите из индивидуального задания таблицы 10.
- Найдите математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X распределенной равномерно на интервале (X,Y) (таб.10 4 и 5 строка).

Таблица 10

№ компьютера	1	2	3	4	5	6	7	8	9	10
a	0	1	2	3	0	2	3	4	0	1
σ	2	1	1	2	3	2	1	3	2	1
X	2	3	4	5	2	3	4	5	2	3
Y	7	8	9	7	8	9	7	8	9	6

Образец выполнения:

1) Введем значения параметров распределения $N(a, \sigma)$, где a - математическое ожидание, σ - среднее квадратическое отклонение случайной величины:

$$a := 0 \quad \sigma := 1$$

2) Определим плотность стандартного нормального распределения и постройте график: $\pi := 3.14$

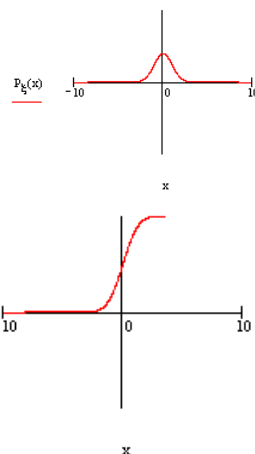
$$P_{\xi}(x) := \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}}$$

3) Определим функцию распределения и постройте график:

$$F_{\xi}(x) := \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^x e^{-\frac{z^2}{2}} dz$$

4) Введем концы интервала (a,b) :

$$a := 2 \quad b := 8$$



5) Определим плотность равномерного распределения:

$$f(x) := \frac{1}{b-a}$$

6) Определим математическое ожидание:

$$M\xi := \int_a^b x \cdot f(x) dx$$

$$M\xi \rightarrow 5$$

7) Определим дисперсию:

$$D\xi := \int_a^b x^2 \cdot f(x) dx - (M\xi)^2$$

$$D\xi \rightarrow 3$$

8) Определим среднее квадратическое отклонение:

$$\sigma := \sqrt{D\xi}$$

$$\sigma \rightarrow \sqrt{3}$$

4. Законы распределения в статистическом пакете STADIA.

Постройте графики плотности распределения вероятностей нормального распределения с параметрами из индивидуального задания.

Образец выполнения:

1). В меню блока Статистические методы (рис.14) щелчком мышью кнопку Т= Вычисление вероятностей или нажмите клавишу Т. На экране появиться меню выбора Функция вероятности распределения в котором необходимо выбрать пункт б= нормальное.

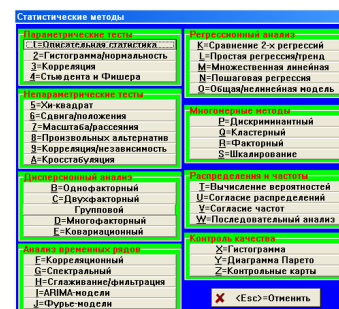
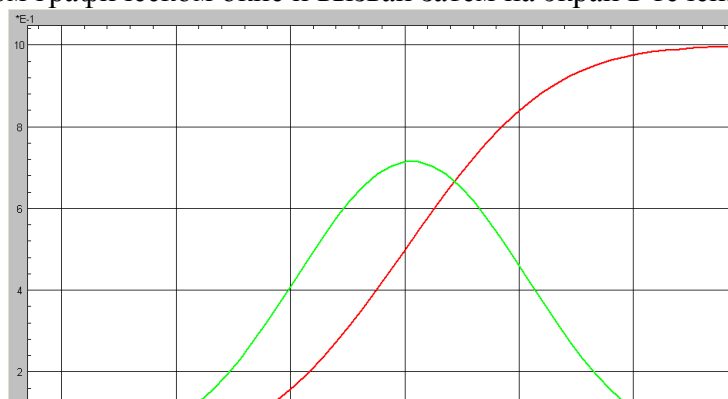


Рис. 14 Меню выбора функции

2). На экране появиться окно ввода параметров нормального распределения, в которое вводится a и σ .

3). Система предложит вычислить значения функции нормального распределения $F(x_1)$ и $F(x_2)$, а так же разности $F(x_1)-F(x_2)$. Точки x_1 и x_2 следует указать в окне ввода параметров. При вводе только одного значения x система выведет указанные величины для пары $-\infty; x$.

4). При нажатии кнопки запроса Утвердить в окне ввода параметров на экран в графическое окно будет выдан график плотности и функции распределения для нормального распределения с заданными параметрами (рис. 15). Полученный график может быть сохранен в отдельном графическом окне и вызван затем на экран в течении всего сеанса работы с пакетом.



Результаты					
ВЫЧИСЛЕНИЕ ВЕРОЯТНОСТЕЙ. файл:					
ВЫЧИСЛЕНИЕ ВЕРОЯТНОСТЕЙ. файл:					
Распределение нормальное: 0, 1					
Среднее=0, Дисперсия=1, Ст.отклонение=1					
x1	x2	F(x1)	F(x2)	F(x1)-F(x2)	
0	0,5	0,5	0,6915	0,1915	
0	0,5	0,5	0,6915	0,1915	
0,5	1	0,6915	0,8414	0,1499	

Рисунок 15 график функции и плотности распределения для нормального распределения.

5. Решите следующую задачу в статистическом пакете STADIA.

Тест для самоконтроля

1. Перечислите виды распределения случайной дискретной величины.
2. Приведите ситуацию (пример), в которой возникает распределение Пуассона.

3. Запишите формулу биномиального распределения вероятностей.
4. Объясните смысл параметров нормального распределения на графике (рис 15).

5. Определите вид распределения для задачи:

Пусть x - число появлений события A в серии из n независимых испытаний. Вероятность появлений события A в k -м испытании равна p_k ($k = 1, 2, \dots, n$).

- а). распределения Пуассона;
- б). Биномиальное распределение;
- в). Нормальное распределение.

6. Определите вид распределения для задачи:

Произведено n независимых бросаний монеты ($n \geq 1$). Случайная величина x - число выпадений герба при этих n бросаниях.

- а). распределения Пуассона;
- б). Биномиальное распределение;
- в). Нормальное распределение.

7. Определите вид распределения для задачи:

В случае нормальной настройки автоматического станка только 1% выпускаемых деталей – дефектные. Предположим, что автомат настроен нормально.

- а). распределения Пуассона;
- б). Биномиальное распределение;
- в). Нормальное распределение.

РАЗДЕЛ №3 ОСНОВНЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Цели изучения: ознакомить с новым типом задач – задач математической статистики; сформировать умения решать задачи с использованием параметров распределения и статистических гипотез; ознакомить с методами сводных характеристик выборки.

Тема №6 Выборочное наблюдение. Ошибки выборки

Изучив эту тему, Вы будете:

Иметь представление:

- о способах сбора и группировки статистических сведений.

Знать:

- виды выборки;
- область использования выборочного метода;

Уметь:

- строить вариационные ряды, таблицы частот, полигон и гистограмму;
- определять необходимый объем выборки;
- моделировать процесс получения случайной величины на ЭВМ;
- определять среднюю и предельную ошибки выборки.

Входной тест

1). Как будет вычисляться формула, содержащая синтаксическую ошибку в ТП Excel?

1. В ячейке будет выведено сообщение об ошибке;
2. Результатом вычисления будет 0;
3. Система не допускает ввода формул с синтаксическими ошибками;
4. В ячейке будет выведено #####.

2). Укажите способ указания на оси ох наименование соответствующего товара при построении диаграммы в ТП Excel.

3). Запиши формулу для вычисления математического ожидания и дисперсии.

4). Где можно изменить заголовок и подписи данных построенной диаграммы в ТП Excel?

1. исходные данные;
2. параметры диаграммы;
3. Тип диаграммы.

Статистические данные представляют собой данные, полученные в результате обследования большого числа объектов или явлений.

Математическая статистика подразделяется на две основные области: описательную и аналитическую статистику. Описательная статистика охватывает методы описания статистических данных, представления их в форме таблиц, распределений и т. п.

Аналитическая статистика или теория статистических выводов ориентирована на обработку данных, полученных в ходе эксперимента, с целью формулировки выводов, имеющих прикладное значение для самых различных областей человеческой деятельности.

Пакет Excel оснащен средствами статистической обработки данных. И хотя Excel существенно уступает специализированным статистическим пакетам обработки данных, тем не менее этот раздел математики представлен в Excel наиболее полно. В него включены основные, наиболее часто используемые статистические процедуры: средства описательной статистики, критерии различия, корреляционные и другие методы, позволяющие проводить необходимый статистический анализ экономических, медико-биологических и иных типов данных.

При рассмотрении применения методов обработки статистических данных ограничимся только простейшими и наиболее часто используемыми методами, реализованными в мастере функций и пакете анализа Excel. Но прежде рассмотрим основные понятия математической статистики.

6.1 Вариационные ряды. Таблица частот. Полигон и гистограмма

Совокупность всех возможных значений случайной величины x в математической статистике называют генеральной совокупностью. Совокупность n возможных значений x : $x_1, x_2, \dots, x_n$, полученных в результате n независимых опытов (наблюдений), называется выборкой или статистическим рядом объема n .

Различные значения случайной величины, содержащиеся в выборке, называются вариантами. Система вариантов $\alpha_1, \alpha_2, \dots, \alpha_m$, расположенных в возрастающем порядке, называются вариационным рядом.

Пусть x – дискретная случайная величина и $\alpha_1, \alpha_2, \dots, \alpha_m$ ее вариационный ряд, полученный по выборке $x_1, x_2, \dots, x_n$. Тогда число $\mu_i = \frac{k_i}{n}$ ($i = 1, 2, \dots, m$), где k_i – количество повторений варианты α_i в выборке объема n , называется частотой этой варианты в данной выборке.

Таблица 11

α_1	α_2	$\dots$	α_m
μ_1	μ_2	$\dots$	μ_m

в первой строке которой расположены варианты, а во второй – соответствующие им частоты, называется таблицей частот или эмпирическим законом распределения дискретной случайной величины x . Сумма всех частот в таблице 11 равна единице.

Если на плоскости в прямоугольной системе координат построим точки (α_i, μ_i) ($i = 1, 2, \dots, m$) и соединим их последовательно отрезками прямых, то получим ломаную линию, которая называется полигоном частот (рис.16)

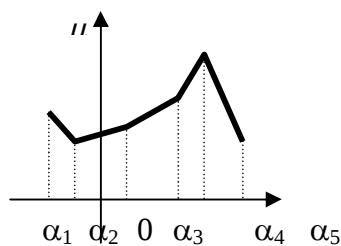


Рис.16

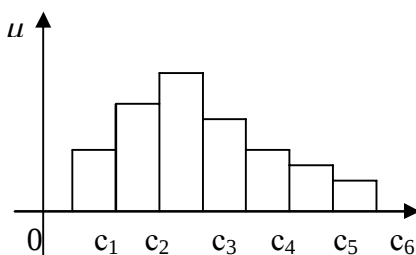


Рис.17

Полигон частот дает приближенное наглядное представление о характере распределения случайной величины x .

Если изучаемая случайная величина x непрерывна, то вместо обычного (дискретного) вариационного ряда составляют интервальный вариационный ряд: находят минимальную и максимальную варианты выборки и весь промежуток между ними разбивают на частичные промежутки. Получается так называемый интервальный вариационный ряд $[c_1; c_2[$, $[c_2; c_3[$, $\dots$, $[c_m; c_{m+1}[$.

Далее по выборке определяют частоту $\tilde{\mu}_i = \frac{\tilde{k}_i}{n}$ ($i = 1, 2, \dots, m$) попадания значений x в i -й интервал. Здесь $\tilde{k}_i$ - количество членов выборки, попавших в i -й интервал. Если при этом некоторое x_k выборки совпадает с граничной точкой между промежутками, то его относят к правому промежутку. В результате получается интервальная таблица частот:

$[c_1; c_2[$	$[c_2; c_3[$	$\dots$	$[c_m; c_{m+1}[$
$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\dots$	$\tilde{\mu}_m$

Интервальная таблица частот графически изображается гистограммой, которая представляет собой ступенчатую линию (рис.17). Основанием i -й ступеньки является i -й частичный интервал, а высота h_i такова, что площадь ступеньки равна $\tilde{\mu}_i$. По построению суммарная площадь всех ступенек гистограммы равна 1.

Пример №1. При изучении некоторой дискретной случайной величины в результате 40 независимых наблюдений получена выборка:

10, 13, 10, 9, 9, 12, 12, 6, 7, 9;

8, 9, 11, 9, 14, 13, 9, 8, 8, 7;

10, 10, 11, 11, 11, 12, 8, 7, 9, 10;

14, 13, 8, 8, 9, 10, 11, 11, 12, 12.

Требуется: а) составить вариационный ряд; б) составить таблицу частот; в) построить полигон.

Выбирая различные варианты из выборки и располагая их в возрастающем порядке, получим вариационный ряд: 6, 7, 8, 9, 10, 11, 12, 13, 14.

б) Для нахождения частот $\mu_i = \frac{k_i}{40}$ предварительно подсчитаем для каждой варианты соответствующие кратности k_i : k_i : 1, 3, 6, 8, 6, 6, 5, 3, 2.

Таблица частот

α_1	6	7	8	9	10	11	12	13	14
μ_1	$\frac{1}{40}$	$\frac{3}{40}$	$\frac{6}{40}$	$\frac{8}{40}$	$\frac{6}{40}$	$\frac{6}{40}$	$\frac{5}{40}$	$\frac{3}{40}$	$\frac{2}{40}$

в) Полигон изображен на рисунке 18.

Пример №2. Для изучения распределения веса новорожденных были собраны данные для 100 детей и составлена интервальная таблица частот.

Интервалы веса (кг)	1,0 – 1,5	1,5 – 2,0	2,0 – 2,5	2,5 – 3,0	3,0 – 3,5	3,5 – 4,0	4,0 – 4,5	4,5 – 5,0
Частота μ_1	0,01	0,02	0,05	0,15	0,35	0,28	0,12	0,02

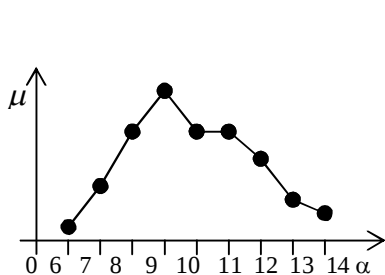


рис. 18 Полигон

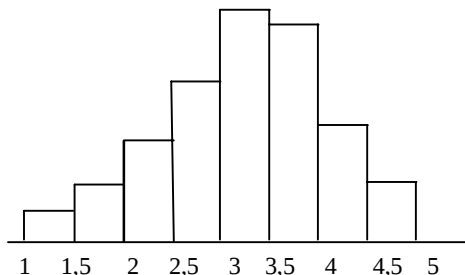


Рис. 19 Гистограмма

Гистограмма полученной интервальной таблицы частот представлена на рисунке 19.

Задание №1

Пятьюдесятью абитуриентами на вступительных экзаменах в МГППК получены следующие количества баллов:

12, 14, 19, 15, 14, 18, 13, 16, 17, 12,
20, 17, 15, 13, 17, 16, 20, 14, 14, 13,
17, 16, 15, 19, 16, 15, 18, 17, 15, 14,
16, 15, 15, 18, 15, 15, 19, 14, 16, 18,
18, 15, 15, 17, 15, 16, 16, 14, 14, 17.

Требуется: а) составить вариационный ряд; б) составить таблицу частот; в) построить полигон частот.

α_1									
μ_1									

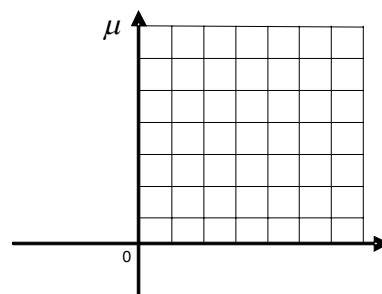
Задание №2

Обследование оплаты труда 60 преподавателей дало следующие результаты (в усл.ед.):

114, 104, 112, 101, 90, 122, 126, 116, 128, 140, 124, 120, 160, 104,
140, 90, 118, 132, 154, 124, 104, 121, 156, 160, 128, 132, 104,82,
130, 114, 142, 122, 160, 98, 116, 98, 132, 142, 116, 126, 108, 121,
102, 104, 122, 96, 122, 138, 124, 123, 108, 121, 102, 104, 122, 96,
122, 138, 124, 123.

а) Составьте интервальную таблицу частот с шириной интервала 10 (у.е.) начиная с 80 (у.е.).

б) Постройте гистограмму.



α_1									
μ_1									

Пример №3

Построить эмпирическое распределение веса студентов в килограммах для следующей выборки:

64,57,63,62,58,61,63,60,60,61,65,62,62,60,64,61,59,59,
63,61,62,58,58,
63,61,59,62,60,60,58,61,60,63,63,58,60,59,60,59,61,62,
62,63,57,61,58,60,64,60,59,61,64,62,59,65.

Воспользуемся процедурой Гистограмма.

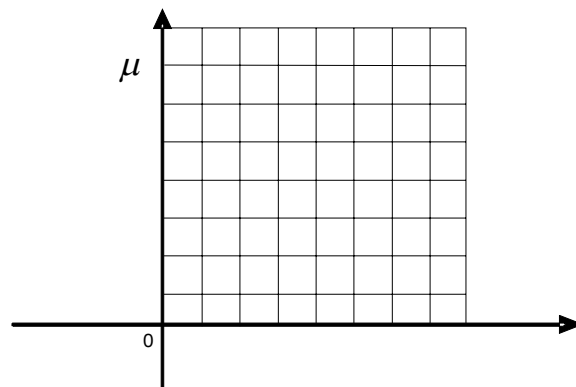
1.В ячейку A1 введем слово Наблюдения, а в диапазон A2:E12 — значения веса студентов.

2.Для вызова процедуры Гистограмма выберем из меню Сервис подпункт Анализ данных и в открывшемся окне в поле Инструменты анализа укажем процедуру Гистограмма.

3.В появившемся окне Гистограмма заполним рабочие поля: во Входной диапазон введем диапазон исследуемых данных (A2:E12); в Выходной диапазон — ссылку на левую верхнюю ячейку выходного диапазона (F1). Установим переключатели в положение Интегральный процент и Вывод графика;

4.После этого нажмем кнопку ОК.

В результате получим таблицу и диаграмму (рис.20).



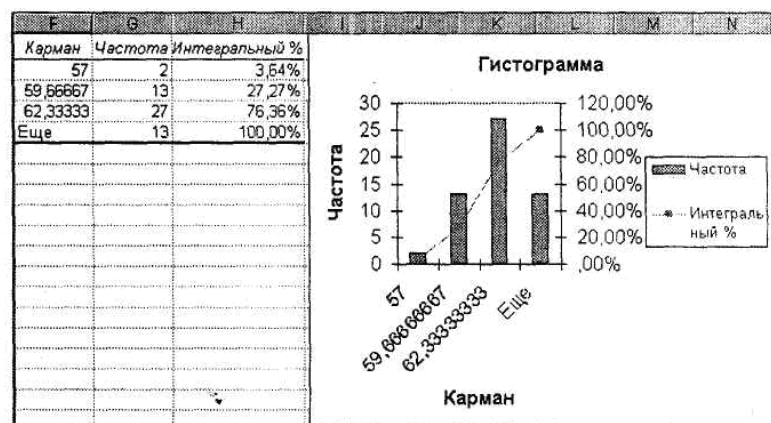


Рис.20 Результаты процедуры Гистограмма пакета Анализа.

6.2 Ошибка выборки

В процессе проведения выборочного наблюдения выделяю 2 вида ошибок: регистрации и репрезентативности. Ошибки регистрации могут иметь случайный или систематический характер. Их можно избежать при правильной организации и проведении наблюдений. Ошибки репрезентативности возникают в силу того, что выборочная совокупность не полностью воспроизводит генеральную. Избежать таких ошибок нельзя, но их можно свести к минимальным значениям, границы которых устанавливаются с достаточно большой точностью.

Расчет ошибок позволяет решить одну из главных проблем организации выборочного наблюдения - оценить репрезентативность (представительность) выборочной совокупности. Различаю среднюю и предельную ошибки выборки. Эти два вида ошибок связаны следующим соотношением:

$$\Delta = t\mu,$$

где Δ - предельная ошибка выборки;

μ - средняя ошибка выборки;

t - коэффициент доверия, определяемый в зависимости от уровня вероятности.

Величина средней ошибки выборки рассчитывается дифференцированно в зависимости от способа отбора и процедуры выборки. Так, при случайном повторном отборе средняя ошибка определяется по формуле из теоремы Чебышева:

$$\mu = \frac{\sigma}{\sqrt{n}},$$

при бесповторном: $\mu = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)},$

где σ^2 - выборочная (или генеральная)² дисперсия

σ - выборочное (или генеральное) среднее

квадратическое отклонение;

n - объем выборочной совокупности;

N - объем генеральной совокупности.

Расчет средней и предельной ошибок выборки позволяет определить возможные пределы, в которых будут находиться характеристики генеральной совокупности. Например, для выборочной средней такие пределы устанавливаются на основе следующих соотношений:

$$\bar{\tilde{x}} - \Delta_{\tilde{x}} \leq \bar{x} \leq \tilde{x} + \Delta_{\tilde{x}},$$

где $\bar{x}$ и $\tilde{x}$ - генеральная и выборочная средние соответственно;

$\Delta_{\tilde{x}}$ - предельная ошибка выборочной средней

по теореме Ляпунова $\Delta_{\tilde{x}} = \pm t \sqrt{\frac{\sigma^2}{n}}$, а вероятность расхождения $|\tilde{x} - \bar{x}|$ равна

$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{t^2}{2}} dt = F(t)$. В частности, при $t=1$ $F(t)=0,683$

$t=2$ $F(t)=0,954$

$t=3$ $F(t)=0,997$.

Покажем практическое применение рассмотренной выше методики на следующих примерах.

Пример № 4

При проверки веса импортируемого груза на таможне методом случайной выборки было отобрано 200 изделий. В результате был установлен средний вес изделия 30 г при среднем квадратическом отклонении 4 г. С вероятностью 0,997 определите пределы, в которых находится средний вес изделий в генеральной совокупности.

Рассчитаем сначала предельную ошибку выборки. Так, при $p = 0,997$, $t=3$

$$\Delta_{\tilde{x}} = t * \frac{\sigma_{\tilde{x}}}{\sqrt{n}} = 3 * \frac{4}{\sqrt{200}} = 0,84$$

Определим пределы генеральной средней:

$30 - 0,84 < \bar{x} < 30 + 0,84$,

или $29,16 < \bar{x} < 30,84$.

Следовательно, с вероятностью 0,997 можно утверждать, что средний вес изделий в генеральной совокупности находится в пределах от 29,16 до 30,84 г.

Задание № 3

В городе проживает 250 тыс. семей. Для определения среднего числа детей в семье была организована 2%-ная случайная бесповторная выборка семей. По ее результатам было получено следующее распределение семей по числу детей:

Число детей в семье	0	1	2	3	4	5
Количество семей	1 000	2 000	12 000	400	200	200

С вероятностью 0,954 найдите пределы, в которых будет находиться среднее число детей в генеральной совокупности.

Вначале на основе имеющегося распределения семей определите выборочные среднюю и дисперсию используя для вычислений таблицу 12 :

$\tilde{x} =$ _____ человека; $\sigma_{\tilde{x}}^2 =$ _____ .

Таблица 12

Число детей в семье x	Количество семей p				
0	1000				
1	2000				
2	1200				
3	400				
4	200				
5	200				
Итого	5000				

Вычислите предельную ошибку выборки (с учетом того, что $p = 0,954$ $t = 2$).

$$\Delta_{\bar{x}} = t \sqrt{\frac{\sigma_{\bar{x}}^2}{2} \left(1 - \frac{n}{N}\right)} = \underline{\hspace{2cm}}$$

Следовательно, пределы генеральной средней:

$$\bar{x} = \tilde{x} \pm \Delta_{\bar{x}} = \underline{\hspace{2cm}}$$

Таким образом, с вероятностью 0,954 можно утверждать, что среднее число детей в семьях города $\underline{\hspace{2cm}}$

Наряду с определением ошибок выборки и пределов для генеральной средней эти же показатели могут быть определены для доли признака. В этом случае особенности расчета связаны с определением дисперсии доли, которая вычисляется так: $\sigma_w^2 = w(1-w)$,

где $w = \frac{m}{n}$ - доля единиц, обладающих данным признаком в выборочной совокупности,

определяемая как отношение количества соответствующих единиц к объему выборки.

Тогда, например, при собственно-случайном повторном отборе для определения предельной ошибки выборки используется следующая формула;

$$\Delta_w = t \sqrt{\frac{\sigma_w^2}{n}} = t \sqrt{\frac{w(1-w)}{n}}.$$

Соответственно при бесповторном отборе

$$\Delta_w = t \sqrt{\frac{\sigma_w^2}{n} \left(1 - \frac{n}{N}\right)} = t \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}.$$

Пределы доли признака в генеральной совокупности p выглядят следующим образом:

$$w - \Delta_w \leq p \leq w + \Delta_w$$

Пример №5. С целью определения средней фактической продолжительности рабочего дня в государственном учреждении с численностью служащих 480 человек в июне 2004 г. была произведена 25%-ная механическая выборка. По результатам наблюдения оказалось, что у 10% обследованных потери времени достигали более 45 мин. в день. С вероятностью 0,683 установите пределы, в которых находится генеральная доля служащих с потерями рабочего времени более 45 мин в день.

Определим объем выборочной совокупности: $n = 480 \cdot 0,25 = 120$ человек.

Выборочная доля w равна по условию 10%. Учитывая, что показатели точности механической и собственно-случайной бес повторной выборки определяются одинаково, а также то, что при $p = 0,683$ $t = 1$, вычислим предельную ошибку выборочной доли:

$$\Delta_w = t \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)} = 1 \sqrt{\frac{0,1(1-0,1)}{120} \left(1 - \frac{120}{480}\right)} = 0,0237 \approx 0,024 \text{ или } 2,4\%.$$

Пределы доли признака в генеральной совокупности:

$$10 - 2,4 \leq p \leq 10 + 2,4,$$

или

$$7,6 \leq p \leq 12,4$$

Таким образом, с вероятностью 0,683 можно утверждать, что доля работников учреждения с потерями рабочего времени более 45 мин. в день находится в пределах от 7,6 до 12,4%,

Ошибки и пределы генеральных характеристик при других способах формирования выборочной совокупности определяются на основе соответствующих формул, отражающих особенности этих видов выборки. Например, в случае типической выборки показателем вариации является средняя из внутригрупповых дисперсий $\overline{\sigma_i^2}$, при серийной выборке - межгрупповая (межсерийная) дисперсия δ^2 и т. д. Кроме того, в последнем случае вместо объема выборочной совокупности n используется показатель числа серий g .

Следовательно, для типической выборки средняя ошибка вычисляется по формулам:

- при отборе, пропорциональном объему типических групп:

$$\mu = \sqrt{\frac{\sigma_i^2}{n}} \text{ (повторный отбор);}$$

$$\mu = \sqrt{\frac{\sigma_i^2}{n} \left(1 - \frac{n}{N}\right)} \text{ (бесповторный отбор);}$$

- при отборе, пропорциональном вариации признака (не пропорциональных объему групп):

$$\mu = \frac{1}{N} \sqrt{\sum \frac{\sigma_i^2 * N_i^2}{n_i}} \text{ (повторный отбор);}$$

$$\mu = \frac{1}{N} \sqrt{\sum \frac{\sigma_i^2 * N_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right)} \text{ (бесповторный отбор),}$$

где N_i - объемы i -й типической группы.

При серийной выборке средняя ошибка определяется следующим образом:

$$\mu = \sqrt{\frac{\delta^2}{r}} \text{ (повторный отбор);}$$

$$\mu = \sqrt{\frac{\delta^2}{r} \left(1 - \frac{r}{R}\right)} \text{ (бесповторный отбор),}$$

где; R - число серий в генеральной совокупности;

$$\sigma^2 = \frac{\sum (\tilde{x}_i - \tilde{\bar{x}})^2}{r} \text{ - межгрупповая (межсерийная) дисперсия;}$$

r - число серий в выборочной совокупности.

Задание № 4. В области, состоящей из 20 районов, проводилось сплошное обследование урожайности на основе отбора серий районов). Выборочные средние по районам составили соответственно 14,5 ц/га; 16; 15,5; 15 и 14 ц/га. С вероятностью 0,954 видите пределы урожайности во всей области.

Рассчитайте общую среднюю:

$$\tilde{\bar{x}} = \frac{14,5 + 16 + 15,5 + 15 + 14}{5} = \underline{\hspace{2cm}} \text{ ц/га.}$$

Межгрупповая (межсерийная) дисперсия

$$\delta^2 = \frac{14,5^2 + 16^2 + 15,5^2 + 15^2 + 14^2}{5} - \tilde{\bar{x}}^2 = \underline{\hspace{2cm}}$$

Определите предельную ошибку серийной бесповторной выборки ($t = 2$, $p = 0,954$):

$$\Delta = \underline{\hspace{1cm}} \sqrt{\frac{\delta^2}{r} \left(1 - \frac{r}{R}\right)} = \underline{\hspace{2cm}}$$

Следовательно, урожайность в области с вероятностью 0,954 :ет находиться в пределах:

$$\underline{\hspace{2cm}} \text{ ц/га} \leq \bar{x} \leq \underline{\hspace{2cm}} \text{ ц/га} .$$

Формулы необходимого объема выборки для различных способов формирования выборочной совокупности могут быть выведены из соответствующих соотношений, используемых в расчете предельных ошибок выборки.

Приведем наиболее часто применяемые на практике выражения необходимого объема выборки:

- собственно-случайная и механическая выборка:

$$n = \frac{t^2 \sigma^2}{\Delta^2} \text{ (повторный отбор);}$$

$$n = \frac{t^2 \sigma^2 N}{\Delta^2 N + t^2 \sigma^2} \text{ (бесповторный отбор);}$$

• типическая выборка:

$$n = \frac{t^2 \overline{\sigma_i^2}}{\Delta^2} \text{ (повторный отбор);}$$

$$n = \frac{t^2 \overline{\sigma_i^2} R}{\Delta^2 R + t^2 \overline{\sigma_i^2}} \text{ (бесповторный отбор);}$$

• серийная выборка:

$$r = \frac{t^2 \cdot \delta^2}{\Delta^2} \text{ (повторный отбор);}$$

$$r = \frac{t^2 \cdot \delta^2 \cdot R}{\Delta^2 R + t^2 \delta^2} \text{ (бесповторный отбор).}$$

При этом в зависимости от целей исследования дисперсии и ошибки выборки могут быть рассчитаны для средней величины или доли признака.

Рассмотрим пример определения необходимого объема выборки.

Пример №6. В 100 туристических агентствах города предполагается провести обследование среднемесячного количества реализованных путевок методом механического отбора. Какова должна быть численность выборки, чтобы с вероятностью 0,683 ошибка не превышала 3 путевок, если по данным пробного обследования дисперсия составляет 225?

Рассчитаем необходимый объем выборки:

$$n = \frac{1^2 \cdot 225 \cdot 100}{3^2 \cdot 100 + 1^2 \cdot 225} = \frac{22500}{1125} = 20 \text{ агентств.}$$

Задание №5. С целью определения доли сотрудников коммерческих банков области в возрасте старше 40 лет предполагается организовать типическую выборку пропорционально численности сотрудников мужского и женского пола с механическим отбором внутри групп. Общее число сотрудников банков составляет 12 тыс. человек, в том числе 7 тыс. мужчин и 5 тыс. женщин.

На основании предыдущих обследований известно, что средняя из внутригрупповых дисперсий составляет 1600. Определите необходимый объем выборки при вероятности 0,997 и ошибке 5%.

Рассчитайте общую численность типической выборки:

$$n = \frac{\quad}{\quad} = \quad \text{человек}$$

Вычислите объем отдельных типических групп:

$$n_1 = \frac{\quad}{\quad} = \quad \text{человек;}$$

$$n_2 = \frac{\quad}{\quad} = \quad \text{человек.}$$

Таким образом, необходимый объем выборочной совокупности сотрудников коммерческих банков составляет 550 человек, в том числе 319 мужчин и 231 женщина.

Практическая работа №4 Ошибки выборки

- По образцу (из примера №3) постройте эмпирические функции распределения (абсолютные и накопленные частоты) успеваемости в вашей группе (для этого в Excel введите в первую колонку фамилии студентов вашей группы, во вторую колонку оценки полученные по дисциплине «Дискретная математика»).
- По данным о крупнейших компаниях Российской Федерации:
 - Произвести отбор (объем выборки равен номеру машины +10);
 - Рассчитать выборочные средние по 2 показателям;

- в.** Вычислить предельные ошибки выборки по этим показателям (уровень вероятности задать самостоятельно);
- г.** Для заданного уровня вероятности определите пределы генеральной средней;
- д.** По выбранным показателям рассчитайте их средние значения для генеральной совокупности;
- е.** Проанализируйте, находятся ли генеральные средние в рассчитанных в пункте 4 пределах;
- ж.** Найдите разности между генеральными и выборочными средними по исследуемым показателям;
- з.** Сравните найденные разности в пункте 7 с предельными ошибками выборки для каждого показателя;
- и.** Определите необходимый объем выборочной совокупности (для заданных в пункте 3 вероятности) таким образом, чтобы ошибка выборки была на 5% меньше полученной для вашей выборки;
- к.** Рассчитайте коэффициенты вариации для выборочной совокупности;
- л.** Вычислите по этим показателям относительные ошибки выборки;
- м.** Сопоставьте относительные ошибки выборки и рассчитанные в пункте 10 коэффициенты вариации. Сформулируйте выводы.

№	Объем Реализации в 2005 г., млрд руб.	Балансовая прибыль в 2005 г., млрд руб.	Прибыль после налогообложения за 2005 г., млрд. руб.	Количество работающих, тыс. чел.
1	14288,6	4999,1	3393,9	160
2	13629,5	525,0	192,7	59
3	13229,3	1364,6	906,5	75
4	10456,2	2186,6	1231,4	47
5	10304,5	1007,3	574,6	50
6	9562,3	1503,4	1089,1	62
7	8106,3	1149,6	779,2	40
8	8057,2	1440,3	1135,4	109
9	7442,2	987,0	841,0	41
10	6480,9	3325,6	1608,1	48
11	5927,4	958,3	582,0	31
12	5312,8	360,2	293,1	37
13	4365,2	305,4	203,2	31
14	4343,6	754,6	554,2	15
15	3985,6	319,2	192,0	34
16	3795,5	1198,3	774,8	12
17	3458,2	130,4	55,1	28
18	3383,6	603,1	277,6	9
19	3340,7	808,4	608,7	23
20	3077,3	399,3	216,1	17
21	3064,1	301,7	178,9	20
22	2854,9	543,4	347,9	26
23	2527,7	960,9	901,8	8
24	2333,6	310,3	198,2	10
25	2265,9	536,0	408,0	3
26	2254,9	306,7	201,0	11
27	2226,6	666,9	547,8	12
28	2196,7	706,5	508,5	17

29	1974,9	866,4	615,5	10
30	1936,2	207,2	86,8	8
31	1932,7	352,4	199,2	13
32	1876,3	218,5	136	25
33	1827,6	290,1	210	12
34	1602,2	157,3	130	23
35	1563,5	91,3	35	14
36	1544,2	270,1	213,2	29
37	1470,2	142,3	125,6	31
38	1462,5	76,3	70,2	35
39	1395,6	138,4	131	18
40	1356,3	107,6	96,3	20
41	1350,2	262,2	198,5	21
42	1341,1	314,3	258,3	32
43	1290,0	190,0	190,2	31
44	1289,1	269,2	269,5	25
45	1285,4	131,2	131,2	14
46	1273,4	200,9	185,2	16
47	1270,1	300,9	156,2	17
48	1267,8	436,2	254,3	18
49	1266,4	243,1	159,4	16
50	1240,1	261,5	196,1	32
51	1228,9	136,1	98,5	38
52	1180,2	166,2	142,3	41
53	1164,9	133,3	98,6	12
54	1128,0	205,0	185,7	35
55	1098,3	136,2	99,9	18

Тест для самоконтроля

- Как называют наблюдаемые значения x_i ?
а) частоты; б) вероятности; варианты.
- Как называют числа наблюдений?
а) частоты; б) вероятности; в) относительные частоты.
- Отклонение выборочных характеристик от соответствующих характеристик генеральной совокупности, возникающее вследствие несплошного характера наблюдения, называется:
а Систематической ошибки репрезентативности;
б Случайной ошибки репрезентативности;
- Определите тип ошибки репрезентативности при следующих условиях отбора: для установления среднего размера вклада населения коммерческих банках производится отбор счетов в соответствии их номерами. Будет ли ошибка выборки случайной?
- Накопленные частоты используются при построении
А).кумуляты
Б). гистограммы
В). Полигона
- Статистическая таблица представляет собой:
А). форму наиболее рационального изложения результатов статистического наблюдения
Б). сведения о чем-нибудь, расположенные по строкам и графам

- В). числовые характеристики, размещенные в колонках таблицы.
6. К статистической таблице можно отнести:
- таблицу умножения
 - опросный лист социологического обследования
 - таблицу, характеризующую численность населения по полу и возрасту.
7. Укажите способ отбора в следующей выборки: при обследовании дневной загрузки продавцов магазина города отбирался каждый десятый среди продовольственных магазинов и каждый пятый среди непродовольственных.
8. Верно ли утверждение. Для определения основных параметров выборки необходимо последовательность вариантов записать в виде вариационного ряда.
9. Чтобы уменьшить ошибку выборки, рассчитанную в условиях механического отбора, можно:
- уменьшить численность выборочной совокупности
 - увеличить численность выборочной совокупности
 - применить серийный отбор
 - применить типический отбор
10. По данным 10%-ного выборочного обследования дисперсия средней заработной платы сотрудников первого туристического агентства 255, а второго-100. Численность сотрудников первого туристического агентства в четыре раза больше, чем второго. Ошибка выборки больше:
- в первом туристическом агентстве
 - во втором туристическом агентстве
 - ошибки одинаковы
 - предсказать результат невозможно
11. На таможенном посту проверенно 36% ручной кладки пассажиров. Ошибка собственно-случайной бесповторной выборки меньше повторной выборки на:
- 10%
 - 20%
 - 80%
 - предсказать результат невозможно

Тема №7 Статистические оценки параметров распределения

Изучив эту тему, Вы будете:

Иметь представление:

- о способах сбора и группировки статистических сведений.

Знать:

- статистические оценки числовых характеристик.

Уметь:

- использовать табличный процессор Excel и математический пакет Mathcad для оценки параметров распределения, выбора объектов из генеральной совокупности, осуществления прогноза случайных величин;
- применять различные виды распределений для нахождения вероятности случайных величин, интервала случайного разброса, доверительного интервала;

Входной тест

- Отклонение выборочных характеристик от соответствующих характеристик генеральной совокупности, возникающее вследствие несплошного характера наблюдения, называется:
 - Систематической ошибки репрезентативности;
 - Случайной ошибки репрезентативности;
- Продолжите предложение. Разность между величиной параметра в генеральной совокупности и его величиной, вычисленной по результатам выборочного наблюдения называется ...

3. Продолжите предложение. Необходимый объем собственно-случайной выборки равен...
4. Какая функция пакета Анализ данных позволяет смоделировать последовательность объектов?
- а. Анализ Фурье;
 - б. Выборка;
 - с. Генерация случайных чисел;

7.1 Выборочные характеристики

Замена теоретической функции распределения $F(x)$ на ее выборочный аналог $F_n(x)$ в определении математического ожидания, дисперсии, стандартного отклонения и т. п. приводят к выборочному среднему, выборочной дисперсии, выборочному стандартному отклонению и т. д. Выборочные характеристики являются оценками соответствующих характеристик генеральной совокупности. Эти оценки должны удовлетворять определенным требованиям. В соответствии с важнейшими требованиями, оценки должны быть:

- несмещенными, то есть стремиться к истинному значению характеристики генеральной совокупности при неограниченном увеличении количества испытаний;
- состоятельными, то есть с ростом размера выборки оценка должна стремиться к значению соответствующего параметра генеральной совокупности с вероятностью, приближающейся к 1;
- эффективными, то есть для выборок равного объема используемая оценка должна иметь минимальную дисперсию.

Среди выборочных характеристик выделяют показатели, относящиеся к центру распределения (меры положения), показатели рассеяния вариант (меры рассеяния) и меры формы распределения. К показателям, характеризующим центр распределения, относят различные виды средних; (арифметическое, геометрическое и т. п.), а также моду и медиану.

Простейшим показателем, характеризующим центр выборки, является мода.

Мода — это элемент выборки с наиболее часто встречающимся значением (наиболее вероятная величина).

Средним значением выборки, или выборочным аналогом математического ожидания, называется величина

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

Иначе говоря, среднее значение — это центр выборки, вокруг которого группируются элементы выборки. При увеличении числа наблюдений среднее приближается к математическому ожиданию. Среднее значение обозначается также буквой M .

Выборочная медиана — это число, которое является серединой выборки, то есть половина чисел имеет значения большие, чем медиана, а половина чисел имеет значения меньшие, чем медиана. Для нахождения медианы обычно выборку ранжируют — располагают элементы в порядке возрастания. Если количество членов ранжированного ряда нечетное, медианой является значение ряда, которое расположено посередине, то есть элемент с номером $(n+1)/2$. Если число членов ряда четное, то медиана равна среднему членов ряда с номерами $n/2$ и $n/2 + 1$.

Основными показателями рассеяния вариант являются интервал, дисперсия выборки, стандартное отклонение и стандартная ошибка.

Интервал (амплитуда, вариационный размах) — это разница между максимальным и минимальным значениями элементов выборки. Интервал является простейшей и наименее надежной мерой вариации или рассеяния элементов в выборке.

Более точно отражают рассеяние показатели, учитывающие не только крайние, но и все значения элементов выборки.

Дисперсией выборки, или выборочным аналогом дисперсии, называется величина

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Дисперсия выборки — это параметр, характеризующий степень разброса элементов выборки относительно среднего значения. Чем больше дисперсия, тем дальше отклоняются значения элементов выборки от среднего значения.

Выборочным стандартным отклонением (среднее квадратичное отклонение) называется величина

$$S = \sqrt{S^2}.$$

Это параметр, также характеризующий степень разброса элементов выборки относительно среднего значения. Чем больше среднее квадратичное отклонение, тем дальше отклоняются значения элементов выборки от среднего значения. Параметр аналогичен дисперсии и используется в тех случаях, когда необходимо, чтобы показатель разброса случайной величины выражался в тех же единицах, что и среднее значение этой случайной величины. Часто выборочное стандартное отклонение обозначают буквой σ (сигма).

Стандартная ошибка или **ошибка среднего** находится из выражения

$$m = \frac{S}{\sqrt{n}}$$

Стандартная ошибка — это параметр, характеризующий степень возможного отклонения среднего значения, полученного на исследуемой ограниченной выборке, от истинного среднего значения, полученного на всей совокупности элементов. С помощью стандартной ошибки задается так называемый доверительный интервал. 95%-ный доверительный интервал, равный $\bar{x} \pm 2m$, обозначает диапазон, в который с вероятностью $p=0,95$ (при достаточно большом числе наблюдений $n > 30$) попадает среднее генеральной совокупности MX .

Выборочной квантилью называется решение уравнения

$$F_n(x) = p.$$

В частности, выборочная медиана есть решение уравнения

$$F_n(x) = 0,5.$$

Показателями, характеризующими форму распределения, являются выборочные эксцесс и асимметрия.

Эксцесс — это степень выраженности «хвостов» распределения, то есть частоты появления удаленных от среднего значений.

Асимметрия — величина, характеризующая несимметричность распределения элементов выборки относительно среднего значения. Принимает значения от -1 до 1. В случае симметричного распределения асимметрия равна 0.

Часто значения асимметрии и эксцесса используют для проверки гипотезы о том, что данные (выборка) принадлежат к определенному теоретическому распределению, в частности, нормальному распределению. Для нормального распределения асимметрия равна нулю, а эксцесс — трем.

7.2 Определение основных статистических характеристик

В результате наблюдений или эксперимента получают наборы данных, называемые выборками. Для проведения их анализа данные подвергаются статистической обработке. Первое, что всегда делается при обработке данных, это вычисление элементарных статистических характеристик выборок (как минимум: среднего, среднеквадратичного отклонения, ошибки среднего) по каждому параметру и по каждой группе. Полезно также вычислить эти характеристики для объединения родственных групп и суммарно по всем данным.

Использование специальных функций.

В мастере функций Excel имеется ряд специальных функций, предназначенных для вычисления выборочных характеристик. Прежде всего, это функции, характеризующие центр распределения.

О Функция СРЗНАЧ вычисляет среднее арифметическое из нескольких массивов (аргументов) чисел. Аргументы *число1, число2, ...* — это от 1 до 30 массивов, для которых вычисляется среднее. Например, если ячейки A1:A7 содержат числа 10, 14, 5, 6, 10, 12 и 13, то сред-

ним арифметическим СРЗНАЧ(A1:A7) является 10.

- о Функция СРГАРМ позволяет получить среднее гармоническое множества данных. Среднее гармоническое — это величина, обратная к среднему арифметическому обратных величин. Например, СРГАРМ(10;14;5;6;10;12;13) равняется 8,317.
- о Функция СРГЕОМ вычисляет среднее геометрическое значений массива положительных чисел. Функцию СРГЕОМ можно использовать для вычисления средних показателей динамического ряда. Например, СРГЕОМ(10;14;5;6;10; 12;13) равняется 9,414.
- о Функция МЕДИАНА позволяет получать медиану заданной выборки. Медиана — это элемент выборки, число элементов выборки со значениями больше которого и меньше которого равно. Например, МЕДИАНА(10;14;5;6;10;12;13) равняется 10.
- о Функция МОДА вычисляет наиболее часто встречающееся значение в выборке. Например, МОДА(10;14;5;6;10;12;13) равняется 10.

К специальным функциям, вычисляющим выборочные характеристики, характеризующие рассеяние вариант, относятся ДИСП, СТАНДОТКЛОН, ПЕРСЕНТИЛЬ.

- о Функция ДИСП позволяет оценить дисперсию по выборочным данным. Например, ДИСП(10;14;5;6;10;12;13) равняется 11,667.
- о Функция СТАНДОТКЛОН вычисляет стандартное отклонение. Например, СТАНДОТКЛОН(10;14;5;6;10;12;13) равняется 3,416.
- о Функция ПЕРСЕНТИЛЬ позволяет получить квантили заданной выборки. Например, если ячейки A1:A7 содержат числа 10,14,5,6,10,12 и 13, то квантилью со значением 0,1 является ПЕРСЕНТИЛЬ(A1:A7;0,1), равная 5,6.

Форму эмпирического распределения позволяют оценить специальные функции ЭКСЦЕСС и СКОС.

- о Функция ЭКСЦЕСС вычисляет оценку эксцесса по выборочным данным. Например, ЭКСЦЕСС(10;14;5;6;10;12;13) равняется -1,169.
- о Функция СКОС позволяет оценить асимметрию выборочного распределения. Например, СКОС(10;14;5;6;10;12;13) равняется -0,527.

Пример №1

Рассматриваются ежемесячные количества реализованных турфирмой путевок за периоды до и после начала активной рекламной компании. Ниже приведены количества реализованных путевок по месяцам.

С рекламой	Без рекламы
162	135
156	126
144	115
137	140
125	121
145	112
151	130

Требуется найти средние значения и стандартные отклонения этих данных.

1. Для проведения статистического анализа прежде всего необходимо ввести данные в рабочую таблицу. Откройте новую рабочую таблицу. Введите в ячейку A1 слово *Реклама*, затем в ячейки A2:A8 — соответствующие значения числа реализованных путевок. В ячейку B1 введите слова *Без рекламы*, а в B2:B8 — значения числа реализованных путевок до начала рекламной компании. Отметим, что рассматриваемые группы данных со статистической точки зрения являются выборками.
2. При статистическом анализе прежде всего необходимо определить характеристики выборки, и важнейшей характеристикой является среднее значение. Определите среднее значение в ячейке (A9).
3. Определите в ячейке B9 среднее значение числа реализованных путевок без активной рекламы.

4. Следующей по важности характеристикой выборки является мера разброса элементов выборки от среднего значения. Такой мерой является среднее квадратичное или стандартное отклонение. Для определения стандартного отклонения в контрольной группе необходимо установить табличный курсор в свободную ячейку (A10). На панели инструментов нажмите кнопку Вставка функции (fx). В появившемся диалоговом окне Мастер функций выберите категорию Статистические и функцию СТАНДОТКЛОН, после чего нажмите кнопку ОК. Появившееся диалоговое окно СТАНДОТКЛОН за серое поле мышью отодвиньте вправо на 1-2см от данных (при нажатой левой кнопке). Указателем мыши введите диапазон данных контрольной группы для определения стандартного отклонения (A2:A8). Нажмите кнопку ОК. В ячейке A10 появится стандартное отклонение выборки — 12,298. Существует правило, согласно которому при отсутствии артефактов данные должны лежать в диапазоне $M \pm 3\sigma$ (в примере $145,7 \pm 36,9$).
5. В ячейке B10 определите стандартное отклонение числа проданных путевок до начала рекламной компании.

Расчета выборочных характеристик можно производить в ТП Excel с помощью инструмента пакета Анализа, в статистическом пакете STADIA. В модульном блоке №2 в учебном элементе №4 были описаны действия, позволяющие определить всю совокупность характеристик случайной величины X.

Л.р.№5 Оценка параметров распределения

1. В массовом производстве производят дискеты. Контролер качества знает, что оборудование для производства дискет плохо отрегулировано и доля брака составляет 8 %. Контракт на покупку дискет допускает 5-процентный коэффициент брака в партии из 1000 дискет.

Создайте модель имитации 20 выборок числа бракованных изделий.

Определите:

- максимальное (минимальное) количество дефектных дискет, которое может оказаться в выборке;
 - среднее число бракованных дискет;
 - относительные характеристики полученных случайных величин.
2. Вы работаете в агентстве по прокату автомобилей, расположенным в Абзаково. По мере приближения зимы вы начинаете отслеживать поступление заявок клиентов на транспорт, снабженный багажниками для перевозки лыж. Через несколько дней после проведения исследования в этой местности выпадает много снега и, как и следовало ожидать, количество вышеупомянутых заявок резко возросло. Итак, на десятый день наблюдения вам нужно узнать, сколько автомобилей, оборудованных багажником для лыж, необходимо приготовить, чтобы полностью удовлетворить спрос в следующий 11 день.

Для решения задачи:

- Сгенерируйте данные за первые 10 дней используя биномиальный закон распределения с вероятностью 0,5 в ячейки A1:A10. А затем активизируйте средство Excel Экспоненциальное сглаживание.
- Введите исходный диапазон (чтобы определить прогноз на следующий день укажите входной интервал на одну строку больше), укажите фактор затухания и метку. Ячейку B1 примените в качестве параметра Выходного интервала.
- Установите флажок Вывод графика и нажмите кнопку Ок.
- Согласно данному сглаженному прогнозу искомая величина находится в ячейки B11. Выделите ячейку с ответом красным цветом.

Фактическое число заявок в одиннадцатый день может упасть на несколько единиц в результате огромного количества причин например изменения погоды.

На графике показан сбалансированный «наплыв» заявок со средним показателем количества заявок в течение всего десятидневного периода.

3. Предположим, вы менеджер отдела обслуживания клиентов фирмы, специализирующейся на разработке программного обеспечения. На днях вы получили от внештатной сотрудницы сообщение по электронной почте, в котором она известила вас, что в последнее время ей постоянно звонят клиенты с жалобами на новые программы вашей фирмы. Вы просите ее зарегистрировать все жалобы, поступающие в течении двух недель и сообщить вам результаты.

Указания к решению задачи:

- Сгенерируйте отчет, который включает ежедневное количество звонков с жалобами на конкретный программный продукт. Эти данные поместите в ячейки A1:A14.
 - Чтобы понять существует ли какая-либо определенная тенденция поступления жалоб, создайте на основе средних данных о полученных звонках скользящее среднее.
5. Определите верхнюю (0,75) и нижнюю (0,25) квартили для выборки результатов измерений роста группы студенток: 164,160,157,166,162,160,161,159,160,163,170, 171

Тест самоконтроля

1. Вариация это
 - А). изменение массовых явлений во времени
 - Б). изменение структуры статистической совокупности в пространстве
 - В). изменение значений признака во времени и пространстве
 - Г). изменение состава совокупности
2. Какой из показателей вариации характеризует абсолютный размер изменчивости признака около средней величины?
 - А). коэффициент вариации
 - Б). дисперсия
 - В). размах вариации
 - Г). среднее квадратическое отклонение
3. Если все значения признака увеличить в 16 раз, то дисперсия:
 - А). не изменится
 - Б). увеличится в 16 раз
 - В). увеличится в 256 раз
 - Г). увеличится в 4 раза
4. Найдите наиболее популярный туристический маршрут из четырех реализуемых фирмой, если за неделю последовательно были реализованы следующие маршруты (приводятся номера маршрутов): 1,3,3,2,1,1,4,4,2,4,1,3,2,4,1,4,4,3,1,2,3,4,1,1, 3.
5. Для константы сглаживания наиболее подходящими являются значения:
 - А. 0,1
 - В. 0,2
 - С. 0,3
6. С помощью какой команды Анализа данных можно определить изменчивость данных?
 - а. экспоненциальное сглаживание
 - б. описательная статистика
 - с. однофакторный дисперсионный анализ

Тема №8 Статистические гипотезы

Изучив эту тему, Вы будете:

Иметь представление:

- о способах сбора и группировки статистических сведений.

Знать:

–

Уметь:

- использовать табличный процессор Excel для проверки статистических гипотез;
- Формы организации учебных занятий модульный урок, практическая работа.

Входной тест

1. Расхождение между расчетными значениями и действительными значениями изучаемых величин называется:
А).ошибкой наблюдения
Б).ошибкой регистрации
В). ошибкой репрезентативности
2. Верно ли утверждение. Для определения основных параметров выборки необходимо последовательность вариант записать в виде вариационного ряда.
3. Какие значения необходимо определить, для того чтобы построить гистограмму относительных частот?
4. Запишите формулу для расчета среднего значения случайной величины.
5. Запишите формулу для расчета стандартного значения случайной величины.

8.1 Проверка статистических гипотез

Помимо описательной статистики, важной областью является также аналитическая статистика. Как уже указывалось в учебном элементе №7, аналитическая статистика или теория статистических выводов ориентирована на обработку данных, полученных в ходе эксперимента, с целью формулировки выводов, имеющих прикладное значение. Здесь решается вопрос, отражают ли наблюдаемые данные объективно существующую реальность. Указанный вопрос решается проверкой соответствующих статистических гипотез. При этом могут выявляться достоверности различий между выборками, взаимосвязи между выборками, влияющие факторы и т. п.

Статистическая гипотеза — это предположение о виде или отдельных параметрах распределения вероятностей, которое подлежит проверке на имеющихся данных.

Проверка статистических гипотез — это процесс формирования решения о возможности принять или отвергнуть утверждение (гипотезу), основанный на информации, полученной из анализа выборки. Методы проверки гипотез называются критериями.

В большинстве случаев рассматривают так называемую нулевую гипотезу (нуль-гипотезу H_0), состоящую в том, что все события произошли случайно, естественным образом. Альтернативная гипотеза (H_1) состоит в том, что события случайным образом произойти не могли, и имело место воздействие некоего фактора.

Обычно нулевая гипотеза формулируется таким образом, чтобы на основании эксперимента или наблюдений ее можно было отвергнуть с заранее заданной вероятностью ошибки α . Эта, заранее заданная вероятность ошибки, называется уровнем значимости.

Уровень значимости — максимальное значение вероятности появления события, при котором событие считается практически невозможным. В статистике наибольшее распространение получил уровень значимости, равный $\alpha=0,05$. Поэтому если вероятность, с которой интересующее событие может произойти случайным образом $p<0,05$, то принято считать это событие маловероятным, и если оно все же произошло, то это не было случайным. В наиболее ответственных случаях, когда требуется особая уверенность в достоверности полученных результатов, надежности выводов, уровень значимости принимают равным $\alpha=0,01$ или даже $\alpha=0,001$.

Величину P , равную $1-\alpha$, называют доверительной вероятностью (уровнем надежности), то есть вероятностью, признанной достаточной для того, чтобы уверенно судить о принятом статистическом решении. Соответственно, в качестве доверительных вероятностей выбирают значения 0,95, 0,99 и 0,999.

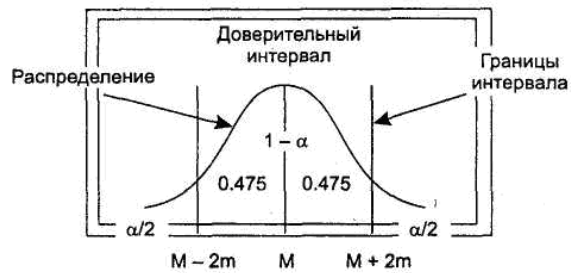


Рис. 21 Геометрическое представление доверительного 95%- интервала

Интервал, в котором с заданной доверительной вероятностью $P=1-\alpha$ находится оцениваемый параметр, называется доверительным интервалом. В соответствии с доверительными вероятностями на практике используются 95%-, 99%-, 99,9%-ные доверительные интервалы. Граничные точки доверительного интервала называют доверительными пределами (рис.21).

Выбор того или иного уровня значимости, выше которого результаты отвергаются как статистически не подтвержденные, или, соответственно, доверительной вероятности, в общем случае является произвольным. Окончательное решение зависит от исследователя, традиций и накопленного практического опыта в данной области исследований.

Анализ однородности выборки. Одним из важных вопросов, возникающих при анализе выборки, является вопрос: относится та или иная варианта к данной статистической совокупности? Решение вопроса не представляет сложности, если распределение в этой совокупности является нормальным. Для этого достаточно использовать правило трех сигм. Согласно этому правилу, в пределах $M \pm 3\sigma$ находится 99,7% всех вариантов. Поэтому, если варианта попадает в этот интервал, то она считается принадлежащей к данной совокупности. Если не попадает, то она может быть отброшена. Хотя этот метод и предполагает нормальность исходного распределения, на практике он успешно работает и может быть использован в большинстве других случаев.

При числе элементов в выборке $n < 30$ способ более точного определения границ доверительного интервала по формуле $[M - t_{n,p}s; M + t_{n,p}s]$

будет показан ниже в примере 1. В формуле M — среднее значение, s — стандартное отклонение, $t_{n,p}$ — табличное значение распределения Стьюдента с числом степеней свободы n и доверительной вероятностью p .

Построение доверительных интервалов для среднего. Еще одной важной задачей, возникающей при анализе одной выборки, является сравнение выборочного среднего арифметического со средним значением генеральной совокупности. Эта задача решается с помощью статистических критериев. При этом выясняется, значимо ли отличие выборочного среднего значения от среднего значения генеральной совокупности, из которой предположительно взята выборка, или наблюдаемое различие является случайным.

Действительно, средние значения, получаемые по выборочным данным, обычно не совпадают с генеральным средним (математическим ожиданием). В связи с этим возникает вопрос: можно ли по результатам выборочной оценки судить о свойствах всей генеральной совокупности?

Поскольку каждую оценку, полученную в отдельной выборке, можно рассматривать как случайную величину, то при увеличении числа выборок распределение отдельных оценок будет принимать характер нормального распределения. Это значит, что в случае средних арифметических значения выборочных средних относительно генерального среднего распределяются по нормальному закону. То есть так же, как относительные отклонения нормально распределенных вариантов от среднего арифметического выборки.

Отсюда, в частности, следует, что 68,3% всех выборочных средних находятся в пределах $\Delta = M \pm m$, где Δ — предельная ошибка выборки, M — среднее выборочное, m — стандартное отклонение среднего значения. Иными словами, имеется вероятность 0,683, что выборочное среднее отличается от генерального не более, чем на $\pm m$. Здесь 0,683 — доверитель-

ная вероятность, $1-0,683=0,317$ — уровень значимости α , $\Delta=M\pm t$ — 68% доверительный интервал.

Для принятой в большинстве исследований доверительной вероятности 0,95, доверительный интервал для средних при достаточно большом числе наблюдений ($n>30$) примерно равен $\pm 2t$ (см. рис. 6.8). При доверительной вероятности 0,99, доверительный интервал составит примерно $\pm 3t$. Для более точного определения границ доверительного интервала можно воспользоваться формулой

$$\left[M - t_{n,p} \frac{s}{\sqrt{n}}; M + t_{n,p} \frac{s}{\sqrt{n}} \right]$$

где M — среднее значение, s — стандартное отклонение, $t_{n,p}$ — табличное значение распределения Стьюдента с числом степеней свободы n и доверительной вероятностью p , n — количество элементов в выборке.

В MS Excel для более точного вычисления границ доверительного интервала и при числе элементов в выборке $n<30$ можно воспользоваться функцией ДОВЕРИТ или процедурой Описательная статистика.

Функция ДОВЕРИТ(*альфа*; *станд_откл*; *размер*) определяет полуширину доверительного интервала и содержит следующие параметры:

- o *Альфа* — уровень значимости, используемый для вычисления доверительной вероятности. Доверительная вероятность равняется $100 \cdot (1 - \text{альфа})\%$ процентам, или, другими словами, *альфа*, равное 0,05, означает 95%-ный уровень доверительной вероятности;
- o *Станд_откл* — стандартное отклонение генеральной совокупности для интервала данных, предполагается известным;
- o *Размер* — это размер выборки.

Пример №1. Найти границы 95%-ного доверительного интервала для среднего значения, если у 25 телефонных аккумуляторов среднее время разряда в режиме ожидания составило 140 часов, а стандартное отклонение — 2,5 часа.

1. Откройте новую рабочую таблицу. Установите табличный курсор в ячейку A1.
2. Для определения границ доверительного интервала необходимо на панели инструментов Стандартная нажать кнопку Вставка функции (fx). В появившемся диалоговом окне Мастер функций выберите категорию Статистические и функцию ДОВЕРИТ, после чего нажмите кнопку ОК.
3. В рабочую область появившегося диалогового окна ДОВЕРИТ с клавиатуры введите условия задачи: Альфа—0,05;Станд_откл—2,5;Размер—25(рис. 22). Нажмите кнопку ОК.
4. В ячейке A1 появится полуширина 95%-ного доверительного интервала для среднего значения выборки — 0,979981. Другими словами, с 95%-ным уровнем надежности можно утверждать, что средняя продолжительность разряда аккумулятора составляет $140 \pm 0,979981$ часа или от 139,02 до 140,98 часа.

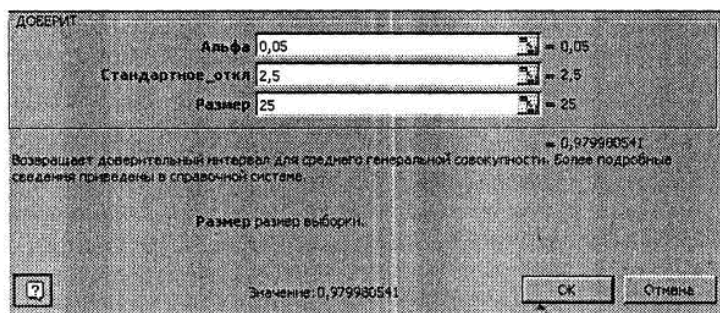


Рис.22 Пример заполнения диалогового окна ДОВЕРИТ

Пример №2

Предположим, ваша компания рассматривает вопрос о приобретении участка для нового магазина, торгующего в розницу. И очень важным критерием при принятии этого решения является место положения магазина, т.е. насколько оно многолюдно. Чтобы выяснить это, вы каждый день в течении двух недель подсчитываете, сколько пешеходов проходит мимо магазина. Результаты этих наблюдений составляют выборку из генеральной совокупности всех возможных дней, в течение которых магазин будет оставаться вашей собственностью.

При этом вы вычисляете *средний показатель* ежедневных результатов наблюдений, и получаете число, например, **403**. *Насколько же точно это среднее значение, полученное на основе выборки, отражает количество людей, которые будут проходить около вашего магазина в любой конкретный день?*

На этот вопрос вам поможет ответить *доверительный интервал среднего значения* (403).

Доверительный интервал-это интервал, с помощью которого возможна оценка с заданной вероятностью неизвестного значения генеральной совокупности. Это неизвестное значение называется доверительным, а его границы- доверительными границами(верхние и нижние границы). В их пределах вы можете обладать неким уровнем уверенности в существовании значения генеральной совокупности.

Чтобы провести такой анализ, выполните следующие действия:

- В ячейку A1 введите Количество прохожих;
- Сгенерируйте количество прохожих с помощью распределения Пуассона. Для этого выберите команду Сервис – Анализ данных –Генерация случайных чисел – распределение Пуассона;
- Укажите в поле Число переменных -1, в поле число случайных чисел – количество дней эксперимента, в поле Параметры – 403;
- Установите переключатель Выходной интервал и укажите ячейку A2;
- Щелкните на кнопку Ок;
- Выберите команду Сервис – Анализ данных – Описательная статистика;
- В поле Выходной интервал введите A1:A15;
- Установите флажок Метки в первой строке и убедитесь, что в поле Уровень надежности установлено значение 95%;
- Установите переключатель Выходной интервал и укажите ячейку C1;
- Щелкните на кнопку Ок;
- В ячейки C5 подсчитайте среднее значение сгенерированных данных;
- В ячейках C6 и C7 подсчитайте верхние и нижние границы доверительного интервала. Для этого введите формулы =СПЗНАЧ(A2:A15)+D3и =СПЗНАЧ(A2:A15)-D3;

Что же означает полученный вами доверительный интервал? Если вы повторите свой эксперимент 100 раз, то получите 100 двухнедельных значений и 100 соответствующих доверительных интервалов. 95 из этих интервалов будут включать средний показатель для генеральной совокупности, т.е. реальное среднее значение генеральной совокупности будет находится между нижней и верхней границами интервала. И только 5 доверительных интервалов не будут охватывать реального среднего числа совокупности.

Проверка соответствия теоретическому распределению. Следующей задачей, возникающей при анализе одной выборки, является оценка меры соответствия (расхождения) полученных эмпирических данных и каких-либо теоретических распределений. Это связано с тем, что в большинстве случаев при решении реальных задач закон распределения и его параметры неизвестны. В то же время применяемые статистические методы в качестве предпосылок часто требуют определенного закона распределения.

Наиболее часто проверяется предположение о нормальном распределении генеральной совокупности, поскольку большинство статистических процедур ориентировано на выборки, полученные из нормально распределенной генеральной совокупности.

Наиболее убедительные результаты дает использование критериев согласия. Критериями согласия называют статистические критерии, предназначенные для проверки согласия опытных данных и теоретической модели. Здесь нулевая гипотеза H_0 представляет собой утверждение о том, что распределение генеральной совокупности, из которой получена выборка, не отличается от нормального. Среди критериев согласия большое распространение получил непараметрический критерий χ^2 (хи-квадрат). Он основан на сравнении эмпирических частот интервалов группировки с теоретическими (ожидаемыми) частотами, рассчитанными по формулам нормального распределения.

Отметим, что сколько-нибудь уверенно о нормальности закона распределения можно судить, если имеется не менее 50 результатов наблюдений. В случаях меньшего числа данных можно говорить только о том, что данные не противоречат нормальному закону, и в этом случае обычно используют графические методы оценки соответствия. При большем числе наблюдений целесообразно совместное использование графических и статистических (например, тест хи-квадрат или аналогичные) методов оценки, естественно дополняющих друг друга.

Использование критерия согласия хи-квадрат. Для применения критерия желательно, чтобы объем выборки $n > 40$, выборочные данные были сгруппированы в интервальный ряд с числом интервалов не менее 7, а в каждом интервале находилось не менее 5 наблюдений (частот).

Отметим, что сравниваться должны именно абсолютные частоты, а не относительные (частоты). При этом, как и любой другой статистический критерий, критерий хи-квадрат не доказывает справедливость нулевой гипотезы (соответствие эмпирического распределения нормальному), а лишь может позволить ее отвергнуть с определенной вероятностью (уровнем значимости).

В MS Excel критерий хи-квадрат реализован в функции ХИ2ТЕСТ. Функция ХИ2ТЕСТ вычисляет вероятность совпадения наблюдаемых (фактических) значений и теоретических (гипотетических) значений. Если вычисленная вероятность ниже уровня значимости (0,05), то нулевая гипотеза отвергается и утверждается, что наблюдаемые значения не соответствуют нормальному закону распределения. Если вычисленная вероятность близка к 1, то можно говорить о высокой степени соответствия экспериментальных данных нормальному закону распределения.

Функция имеет следующие параметры: ХИ2ТЕСТ(*фактический_интервал*; *ожидаемый_интервал*). Здесь:

- о *фактический_интервал* — это интервал данных, которые содержат наблюдения, подлежащие сравнению с ожидаемыми значениями;
- о *ожидаемый_интервал* — это интервал данных, который содержит теоретические (ожидаемые) значения для соответствующих наблюдаемых.

Пример №3. Проверить соответствие выборочных данных. (64,57,63,62,58,61,63,60,60,61,65,62,62,60,64,61,59,59,63,61,62,58,58,63,61,59,62,60,60,58,61,60,63,63,58,60,59,60,59,61,62,62,63,57,61,58,60,64,60,59,61,64,62,59,65) нормальному закону распределения.

1. Повторите пункты 1-4 решения примера 1.
2. Найдите теоретические частоты нормального распределения. Для этого предварительно необходимо найти среднее значение и стандартное отклонение выборки.

В ячейке I13 с помощью функции СРЗНАЧ найдите среднее значение для данных из диапазона A2:E12 (60,855). В ячейке J13 с помощью функции СТАНДОТКЛОН найдите стандартное отклонение для этих же данных (2,05). В ячейки K1 и K2 введите название столбца — *Теоретические частоты*. Затем с помощью функции НОРМРАСП найдите теоретические частоты. Установите курсор в ячейку K4, вызовите указанную функцию и запол-

ните ее рабочие поля: x — G4; Среднее — $I\$13$; Стандартное_откл — $J\$13$; Интегральный — 0. Получим в ячейке K4 0,033. Далее протягиванием скопируйте содержимое ячейки K4 в диапазон ячеек K5:K12. Затем в ячейки L1 и L2 введите название нового столбца — *Теоретические частоты*. Установите курсор в ячейку L4 и введите формулу $=H\$13*K4$. Далее протягиванием скопируйте содержимое ячейки L4 в диапазон ячеек L5:L12.

3. С помощью функции ХИ2ТЕСТ определите соответствие данных нормальному закону распределения. Для этого установите табличный курсор в свободную ячейку L13. На панели инструментов Стандартная нажмите кнопку Вставка функции (fx). В появившемся диалоговом окне Мастер функций выберите категорию Статистические и функцию ХИ2ТЕСТ, после чего нажмите кнопку ОК. Указателем мыши в рабочие поля введите фактический H4:H12 и ожидаемые L4:L12 диапазоны частот (рис.23). Нажмите кнопку ОК. В ячейке L13 появится значение вероятности того, что выборочные данные соответствуют нормальному закону распределения — 0,9842.

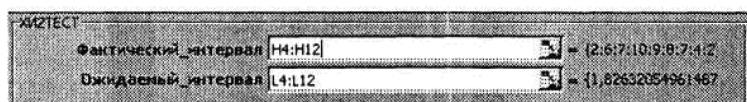


Рис.23 Ввод данных в диалоговом окне расчета функции ХИ2ТЕСТ

4. Поскольку полученная вероятность соответствия экспериментальных данных $p=0,98$ много больше, чем уровень значимости $\alpha = 0,05$, то можно утверждать, что нулевая гипотеза не может быть отвергнута и, следовательно, данные не противоречат нормальному закону распределения. Более того, поскольку полученная вероятность $p=0,98$ близка к 1, можно говорить о высокой степени вероятности того, что экспериментальные данные соответствуют нормальному закону.

Л.р.№6 Проверка гипотезы. Применение критерия согласия для проверки нормального распределения генеральной совокупности

Решите следующие задачи с помощью ТП Excel и статистического пакета STADIA.

1. Менеджер по закупкам отдела «Книга - почтой» отслеживал ежедневные заказы на техническую литературу и регистрировал объем продаж в течении 2 месяцев (61 день). Среднее число заказов в день составило $38+a$ (где a - количество букв в вашем полном имени). Требуется определить:
 1. В каком интервале лежит реальное среднее число заказов технической литературы;
 2. Максимальный и минимальный размер заказа;
 3. Насколько отличаются друг от друга размеры заказов на техническую литературу.
2. Определите, лежит ли значение 19 внутри границ 95%.-ного доверительного интервала выборки 2,3,5,7,4,9,6,4,9,10,4,7,19.
3. Опросите студентов вашей подгруппы узнав вес каждого студента в кг. Найдите соответствие полученных вами экспериментальных данных нормальному закону распределения.

Тест для самоконтроля

- 4) Продолжи утверждение. Нулевой называют гипотезу, которая ...
- 5) Какую роль играет уровень значимости критерия при проверки гипотез?
- 6) Выберите из перечисленного формулу для определения выборочного стандартного отклонения:

$$A) s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad B) s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad B) s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i + \bar{x})^2}$$

- 7) При каком условии выдвинутая гипотеза принимается?
 - А) если определяемое значение критерия меньше или равно критического значения;
 - Б) если определяемое значение критерия больше или равно критического значения;
 - В) если определяемое значение критерия равно критическому значению.

- 8) В качестве одного из тестов для оценки уровня физической подготовки студентов 1-го курса одного из колледжей были выбраны прыжки в длину с места. Результаты контрольной группы студентов в количестве 11 человек следующие (в см): 212, 223, 225, 208, 225, 212, 241, 208, 212, 208, 208. Проверить гипотезу: Можно ли по данным результатам судить о хорошей подготовке студентов, если среднее значение длины прыжка – 225 см?

РАЗДЕЛ №4 МЕТОД СТАТИСТИЧЕСКИХ ИСПЫТАНИЙ

Цели изучения: ознакомить с методами обработки результатов наблюдений массовых явлений. Сформировать навыки моделирования случайных величин.

Тема №9 Моделирование случайных величин

Изучив эту тему, Вы будете:

Иметь представление:

- о технологии создания имитаций случайных величин на ЭВМ.

Знать:

- идею метода статистических испытаний;
- общий алгоритм моделирования;

Уметь:

- моделировать последовательность случайных испытаний;

Входной тест

1. Задан закон распределения:

X	6	3	1
P	0,2	0,3	0,5

математическое ожидание дискретной случайной величины равно:

- а) 6
- б) 7,3
- в) 2,6

2. Запишите формулу для расчета среднего квадратического отклонения.

3. Вычислите определенный интеграл $\int_1^3 (x+1)dx$.

9.1 Общая идея метода статистических испытаний

Во всех рассмотренных вероятностных задачах МБ. №1-3 удавалось установить формальную зависимость конечного результата от исходных данных, то есть получить аналитическое решение задачи. Если этого сделать нельзя, то используют метод статистических испытаний.

Основная идея метода состоит в следующем: вместо аналитического решения задачи либо проводятся эксперименты, испытания, непосредственно рассматриваемые в задаче, либо эти испытания заменяются другими, имеющими с исходными одинаковую вероятностную структуру, или, иначе говоря рассматриваемые в задаче случайные явления имитируют, моделируют другими случайными явлениями.

Метод статистических испытаний применяют для решения не только тех задач, в которых в явном виде имеются случайные явления, но также и для решения многих математических задач, не содержащих таких явлений. В этом случае искусственно подбирают такое случайное явление, характеристики которого связаны с результатом решения исходной задачи. Для определения числовых значений этих характеристик используется метод статистических испытаний.

Так как достаточно высокая точность решения при использовании метода статистических испытаний гарантируется, как правило, только при проведении большого числа испытаний, этот метод практически можно реализовать только на компьютере.

9.2 Последовательность зависимых испытаний

Проводится последовательность зависимых испытаний. В результате каждого испытания может произойти одно из противоположных испытаний А и В.

Моделирование этой последовательности испытаний осуществляется так:

1. получаем значение r_1 случайной величины R. Если $r_1 < P_1(A)$, где $P_1(A)$ - вероятность наступления события А в 1-м испытании, то считаем, что в 1-м испытании произошло событие А. Если $r_1 \geq P_1(A)$, то фиксируем появление события В. Допустим, что в первом испытании появилось событие А;
2. получаем следующее значение r_2 . Если $r_2 < P_2(A/A)$ где $P_2(A/A)$ - условная вероятность появления во 2-м испытании события А, то фиксируем появление во 2-м испытании события А при условии, что в первом испытании произошло событие А, то фиксируем появление во 2-м испытании событие А; если $r_2 \geq P_2(A/A)$, считаем, что произошло событие В. Допустим, что произошло В;
3. получаем следующее значение r_3 . Если $r_3 < P_3(A/AB)$, где $P_3(A/AB)$ - вероятность появления в 3-м испытании события А при условии наступления в 1-м и 2-м испытаниях событий А и В, то считаем, что в 3-м испытании появилось событие А; в противном случае событие В и т.д.

Этот алгоритм моделирования зависимых испытаний легко обобщить на случай, когда результатом каждого испытания может быть появление не только двух, а и большего числа событий.

9.3 Моделирование дискретной случайной величины. Общий алгоритм моделирования

Если случайная величина дискретная, то ее моделирование (получение последовательности ее значений) можно свести к моделированию независимых испытаний. Действительно, пусть имеет место следующий ряд распределения:

х	x_1	x_2	...	x_n
р	p_1	p_2	...	p_n

Обозначим A_i событие, состоящее в том, что случайная величина X примет значение x_i . Тогда нахождение значения, принятого случайной величиной X в результате испытания, сводится к определению того, какое из событий $A_1, A_2, \dots, A_n$ несовместны, образуют полную группу и вероятность появления каждого из них не изменяется от испытания к испытанию, то для определения последовательности значений, принятых случайной величиной X , можно использовать процедуру моделирования последовательности независимых испытаний.

Помимо рассмотренного выше общего алгоритма моделирования случайной дискретной величины для многих законов распределения существуют специальные алгоритмы. Рассмотрим специальные алгоритмы моделирования случайных величин с биномиальным распределением и распределением Пуассона.

9.4 Моделирование случайной величины с биномиальным распределением

Напомним, что в соответствии с биномиальным распределением вероятность $P_n(m)$ того, что определенное событие появиться m раз в n зависимых испытаниях:

$$P_n(m) = C_n^m p^m (1-p)^{n-m}.$$

Где p - вероятность появления события в каждом отдельно взятом испытании.

Введем случайную величину X_i - число появлений события в i -м испытании, $i=1, 2, \dots, n$. Очевидно, что эта величина может принимать только два значения: либо 1 с вероятностью p , либо 0 с вероятностью $(1-p)$, т.е. ряд распределения:

X_i	0	1
P	p	$1-p$

Тогда случайное число m появлений события в n испытаниях:

$$m = X_1 + X_2 + \dots + X_n.$$

Исходя из соотношения и распределения, определение значения случайной величины m сводится к следующей процедуре:

1. получают последовательность значений $r_1, r_2, \dots, r_n$ случайной величины R ;
2. для каждого числа $r_i, i=1, 2, \dots, n$. Проверяют, выполняется ли неравенство $r_i < p$. Если неравенство выполняется, то полагают $X_i=1$; в противном случае считают $X_i=0$;
3. находят сумму значений n случайных величин X_i (это и будет значение случайной величины m). Повторяя эту процедуру, получают последовательность значений $m_1, m_2, \dots$ случайной величины с биномиальным законом распределения.

9.5 Моделирование случайной величины с равномерным распределением на отрезке $[a, b]$

Пусть случайная величина X имеет равномерное распределение на отрезке $[a, b]$. Тогда для $X \in [a, b]$ функция распределения:

$$F_x(x) = \frac{x-a}{b-a}.$$

Имеем: $\frac{x-a}{b-a} = r.$

Отсюда:

$$x = a + r(b-a).$$

Последовательность значений $r_1, r_2, \dots$ случайной величины R соответствует последовательность значений $x_1=a+r_1(b-a), x_2=a+r_2(b-a), \dots$ величины X , равномерно распределенной на отрезке $[a, b]$.

9.6 Метод Монте - Карло

Сущность метода Монте- Карло состоит в следующем: требуется найти значение a некоторой изучаемой величины. Для этого выбирают такую случайную величину X , математическое ожидание которой равно a : $M(X)=a$.

Практически же поступают так: производят n испытаний, в результате которых получают n возможных значений X ; вычисляют их среднее арифметическое

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

и принимают $\bar{x}$ в качестве оценки (приближённого значения) a^* искомого числа a :

$$a \approx a^* = \bar{x}.$$

Поскольку метод Монте-Карло требует проведения большого числа испытаний, его часто называют методом статистических испытаний. Теория этого метода указывает, как

наиболее целесообразно выбрать случайную величину X , как найти её возможные значения. В частности, разрабатываются способы уменьшения дисперсии используемых случайных величин, в результате чего уменьшается ошибка, допускаемая при замене искомого математического ожидания a его оценкой a^* .

9.7 Оценка погрешности метода Монте-Карло

Пусть для получения оценки a^* математического ожидания a случайной величины X было произведено n независимых испытаний (разыграно n возможных значений X) и по ним была найдена выборочная средняя $\bar{X}$, которая принята в качестве искомой оценки: $a^* = \bar{X}$. Ясно, что если повторить опыт, то будут получены другие возможные значения X , следовательно, другая средняя, а значит, и другая оценка a^* . Уже отсюда следует, что получить точную оценку математического ожидания невозможно. Естественно возникает вопрос о величине допускаемой ошибки. Ограничимся отысканием лишь верхней границы d допускаемой ошибки с заданной вероятностью (надёжностью) g : $P(|\bar{X} - a| \leq \delta) = g$.

Интересующая нас верхняя грань ошибки d есть не что иное, как «точность оценки» математического ожидания по выборочной средней при помощи доверительных интервалов. Рассмотрим следующие три случая.

Случайная величина X распределена нормально и её среднее квадратичное отклонение d известно.

В этом случае с надёжностью g верхняя граница ошибки

$$\delta = \frac{t \cdot \sigma}{\sqrt{n}}, (*)$$

где n – число испытаний (разыгранных значений X); t – значение аргумента функции Лапласа, при котором $\Phi(t) = \frac{g}{2}$, σ – известное среднее квадратичное отклонение X .

Случайная величина X распределена нормально, причём её среднее квадратическое отклонение s неизвестно.

В этом случае с надёжностью g верхняя граница ошибки $\delta = \frac{t_g \cdot s}{\sqrt{n}}$, (**)

где n – число испытаний; s – «исправленное» среднее квадратическое отклонение,

Случайная величина X распределена по закону, отличному от нормального.

В этом случае при достаточно большом числе испытаний ($n > 30$) с надёжностью, приближённо равной g , верхняя граница ошибки может быть вычислена по формуле (*), если среднее квадратическое отклонение s случайной величины X известно; если же s неизвестно, то можно подставить в формулу (*) его оценку s – «исправленное» среднее квадратическое отклонение либо воспользоваться формулой (**). Заметим, что чем больше n , тем меньше различие между результатами, которые дают обе формулы. Это объясняется тем, что при $n \rightarrow \infty$ распределение Стьюдента стремится к нормальному.

Из изложенного следует, что метод Монте-Карло тесно связан с задачами теории вероятностей, математической статистики и вычислительной математики. В связи с задачей мо-

делирования случайных величин (в особенности равномерно распределённых) существенную роль играют также методы теории чисел.

Среди других вычислительных методов, метод Монте-Карло выделяется своей простотой и общностью. Медленная сходимость является существенным недостатком метода, однако, могут быть указаны его модификации, которые обеспечивают высокий порядок сходимости при определённых предположениях. Правда, вычислительная процедура при этом усложняется и приближается по своей сложности к другим процедурам вычислительной математики. Сходимость метода Монте-Карло является сходимостью по вероятности. Это обстоятельство вряд ли следует относить к числу его недостатков, ибо вероятностные методы в достаточной мере оправдывают себя в практических приложениях. Что же касается задач, имеющих вероятностное описание, то сходимость по вероятности является даже в какой-то мере естественной при их исследовании.

Практическая работа №7 Имитация случайных величин на ЭВМ

$$I = \int_a^b f(x) dx$$

1. Вычислить определенный интеграл применяя метод “Монте-Карло” по формуле

$$I = \frac{1}{n} \cdot \sum_{i=1}^n \frac{f(x_i)}{g(x_i)}, \text{ где } n - \text{число испытаний}; g(x) - \text{плотность распределения “вспомогательной” случайной величины } X, \text{ причем } \int_a^b g(x) dx = 1, \text{ в программе } g(x) = 1/(b-a)$$

Для этого написать программу на языке программирования, которая позволяет вычислить

определенный интеграл: $\int_1^3 (x+1) dx = 6$; $\int_0^3 x^2 dx = 9$.

Требуется ввести промежуток интегрирования и количество испытаний, интегрируемая функция можно задать в программе.

Пример результатов работы программы

Функция	k	N=10	N=100	N=500	N=1000
F(x)=1+x	2	5.737	5.9702	6.02	5.99
F(x)=x*x	3	9.6775	8.528	8.7463	8.937

2. В Mathade сгенерируйте выборку объема n значений случайной величины x имеющей нормальное распределение N(a,b) с помощью функции `gnorm(n,a,b)`, значением которой является вектор, содержащий n выборочных значений нормально распределенной величины с математическим ожиданием a и средним квадратическим отклонением b.

`a := 150 b := 10`

`ξ := rnorm(n,a,b)`

	0
0	145.61
1	143.206
2	145.267
3	140.485
4	133.143
5	150.435
6	148.794
7	155.564
8	171.918
9	158.087
10	159.851
11	158.622
12	159.156
13	156.73
14	139.557
15	150.691

$\xi =$

4) Упорядочите выборку по возрастанию (вариационный ряд):

$\xi := \text{sort}(\xi)$

	0
0	120.991
1	121.596
2	122.741
3	127.059
4	128.429
5	129.688
6	130.834
7	131.784
8	132.828
9	133.062
10	133.143
11	133.692
12	133.852
13	134.241
14	134.377
15	135.371

$\xi =$

5) Вычислите максимальное, минимальное значения и размах выборки:

$\xi_{\max} := \max(\xi) \quad \xi_{\min} := \min(\xi) \quad R := \xi_{\max} - \xi_{\min}$

$\xi_{\max} = 180.482 \quad \xi_{\min} = 120.991 \quad R = 59.49$

6) Определите число интервалов группировки m и их длину ?:

$m := 10 \quad \Delta := \frac{R}{m}$

Δ

7) Определите середины интервалов группировки:

$j := 1 \dots m$

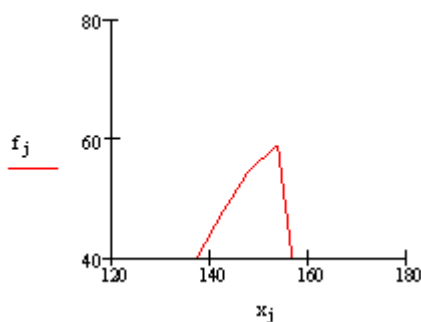
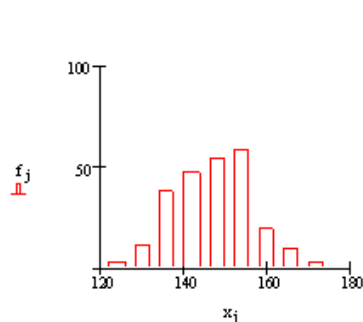
$x_j := \xi_{\min} + \frac{\Delta}{2} \cdot (2 \cdot j - 1)$

	0
0	0
1	123.966
2	129.915
3	135.864
4	141.813
5	147.762
6	153.711
7	159.66
8	165.609
9	171.558
10	177.507

x =

8) Постройте гистограмму и полигон частот с помощью функции $\text{hist}(x, \xi)$, где x - вектор середин интервалов группировки, ξ - выборка. При построении гистограммы в окне настройки изображения графиков пометьте Crossed в пункте X-Y-Axes и установите тип линии bar в пунктире Traces.

$f := \text{hist}(x, \xi)$



9) Повторите все вычисления для $m=20$ и $m=100$, изменяя соответствующие значение в пункте 6).

10) Повторите все вычисления выборки объема $n=300$ и $m=50$, изменяя соответствующие значение в пунктах 2) и 6).

3. Составьте программу моделирование случайной величины для студентов номер машины которых четное число - с равномерным распределением на отрезке $[a,b]$; для студентов номер машины которых нечетное число - с биномиальным распределением на отрезке $[a,b]$;

Тест для самоконтроля

1. Выберите правильный ответ. С надёжностью g верхняя граница ошибки выборки равна:

a). $\delta = \frac{t_\gamma \cdot s}{\sqrt{n}}$

б). $\delta = \frac{t_\gamma \cdot a}{\sqrt{n}}$

$$в). \delta = \frac{t_{\gamma} \cdot a}{\sqrt{\sigma}}$$

$$г). \delta = \frac{t_{\gamma} \cdot s}{\sqrt{\sigma}}$$

2. Установите соответствие вида распределения и встроенной функции в Mathcad:

- | | |
|--|-------------------|
| 1. генератор нормального распределения | а). dbinom(k,n,p) |
| 2. генератор биномиального распределения | |
| 3. функция нормального распределения | б). dnorm(k,n,p) |
| 4. функция биномиального распределения | в). rbinom(k,n,p) |

г). rnorm(k,n,p)

3. Составьте алгоритм, для моделирования дискретной случайной величины.
4. Составьте алгоритм, для моделирования случайной величины с биномиальным распределением.

Тема №10. Статистический метод управления качеством

Изучив эту тему, Вы будете:

Иметь представление:

- о технологии создания имитаций случайных величин на ЭВМ.

знать:

- способ расчета контрольных карт качества в статистическом методе управления качеством в MathCAD и Excel;
- технологию решения задач случайных явлений с помощью ЭВМ.

уметь:

- строить контрольные карты качества с помощью математического пакета Mathcad.

Входной тест

7. Какой из показателей вариации характеризует абсолютный размер изменчивости признака около средней величины?
- А). коэффициент вариации
Б). дисперсия
В). размах вариации
Г). среднее квадратическое отклонение
8. Вычислите δ , если известно $M(x)=5$, $M(x^2) = 34$
9. Определите тип ошибки репрезентативности при следующих условиях отбора: для установления среднего размера вклада населения коммерческих банках производится отбор счетов в соответствии их номерами. Будет ли ошибка выборки случайной?
10. Как называют числа наблюдений?
- а) частоты; б) вероятности; в) относительные частоты.

10.1 Основные понятия

При организации любого производственного процесса возникает задача установки пределов характеристик изделия, в рамках которых произведенная продукция удовлетворяет своему предназначению. Вообще говоря, существует два "врага" качества продукции: (1) отклонения от плановых спецификаций и (2) слишком большой разброс реальных характеристик изделий (относительно плановых спецификаций). На ранних стадиях отладки производственного процесса для оптимизации этих двух показателей качества часто используются методы планирования эксперимента. Методы, содержащиеся в модуле "Контроль качества", предназначены для построения процедур контроля качества продукции в процессе ее производства, т.е. текущего контроля качества. Общий подход к текущему контролю качества достаточно прост. В процессе производства проводятся выборки изделий заданного объема. После этого на специально разлинованной бумаге строятся диаграммы изменчивости выборочных значений плановых спецификаций в этих выборках и рассматривается степень их близости к заданным значениям. Если диаграммы обнаруживают наличие выборочных значений или оказывается, что выборочные значения находятся вне заданных пределов, то считается, что процесс вышел из-под контроля, и предпринимаются необходимые действия для того, чтобы найти причину его разладки. Иногда такие специально разлинованные бумаги называют контрольными картами Шуэрта (в честь W. A. Shewhart, который общепризнанно считается первым, применившим на практике описываемые здесь методы анализа).

Классификация типов контрольных карт часто осуществляется согласно типам величин, которые выбраны для отслеживания характеристик качества. Так, различают контрольные карты для непрерывных *переменных* и контрольные карты по *альтернативному признаку*. В частности, для контроля по непрерывному признаку обычно строятся следующие контрольные карты:

- **Х-карта.** На эту контрольную карту наносятся значения выборочных средних для того, чтобы контролировать отклонение от среднего значения непрерывной переменной (например, диаметров поршневых колец, прочности материала и т.д.).
- **Р-карта.** Для контроля за степенью изменчивости непрерывной величины в контрольной карте этого типа строятся значения размахов выборок.
- **S-карта.** Для контроля за степенью изменчивости непрерывной переменной в контрольной карте данного типа рассматриваются значения выборочных стандартных отклонений.
- **S**2-карта.** В контрольной карте данного типа для контроля изменчивости строится график выборочных дисперсий.

Для контроля качества продукции по альтернативному признаку обычно применяются следующие типы контрольных карт:

- **С-карта.** В таких контрольных картах строится график числа дефектов (в партии, в день, на один станок и т.п.). При использовании карты этого типа делается предположение, что дефекты контролируемой характеристики продукции встречаются сравнительно редко, при этом контрольные пределы для данного типа карт рассчитываются на основе свойств распределения Пуассона (распределения редких событий).
- **U-карта.** В карте данного типа строится график относительной частоты дефектов, то есть отношения числа обнаруженных дефектов к n - числу проверенных единиц продукции (здесь n обозначает, например, число футов длины трубы, объем партии изделий). В отличие от С-карты, для построения карты данного типа не требуется постоянство числа единиц проверяемых изделий, поэтому ее можно использовать при анализе партий различного объема.
- **Np-карта.** В контрольных картах этого типа строится график для числа дефектов (в партии, в день, на станок), как и в случае С-карты. Однако, контрольные пределы этой карты рассчитываются на основе биномиального распределения, а не распределения редких событий Пуассона. Поэтому данный тип карт должен использоваться в том случае, когда обнаружение дефекта не является редким событием (например, когда обнаружение дефекта проис-

ходит более чем у 5% проверенных единиц продукции). Этой картой можно воспользоваться, например, при контроле числа единиц продукции, имеющих небольшой брак.

- **Р-карта.** В картах данного типа строится график процента обнаруженных дефектных изделий (в расчете на партию, в день, на станок и т.д.). График строится так же, как и в случае U-карты. Однако контрольные пределы для данной карты находятся на основе биномиального распределения (для долей), а не распределения редких событий. Поэтому Р-карта наиболее часто используется, когда появление дефекта нельзя считать редким событием (если, например, ожидается, что дефекты будут присутствовать в более чем 5% общего числа произведенных единиц продукции).

Все перечисленные выше типы карт допускают возможность построения кратких карт для производственных серий и контрольных карт для нескольких процессов

Контрольные карты качества (ККК) представляют собой вспомогательное средство для контроля и управления процессами в промышленном серийном производстве в отношении качества промежуточных и конечных продуктов. Для того чтобы избежать появления брака, в некоторые моменты времени (равноотстоящие моменты в процессе производства) берутся выборки продукции, оцениваются, и результаты этой оценки графически фиксируются на ККК. ККК по Шеворту (на которых не указаны границы допуска) характеризуются своими верхними и нижними предупреждающими границами и границами вмешательства (ВГВ, НГВ, ВПГ и НПП). Средняя линия карты — это математическое ожидание контролируемой функции. Границы ККК представляют собой границы 99%-ного (границы вмешательства) и 95%-ного (предупреждающие границы) интервалов разброса.

Пример №1 (контрольные карты качества по Шеворту для математического ожидания и дисперсии; $\bar{x}$ - s - карты)

Из предварительных проверок нам известно, что (насыпной) вес упаковки спагетти распределен нормально. При этом математическое ожидание составляет 500 г, а стандартное отклонение σ - 1 г. Мы должны контролировать процесс расфасовки при помощи $\bar{x}$ - s - карты, используя выборки объема $n=7$. Какими будут границы такой двойной карты? Как она будет выглядеть в масштабном представлении?

Кроме того, нас интересует, как отреагирует контрольная карта средних значений на смещение среднего значения на $\mu_t=501.5$, т.е. какой будет вероятность вмешательства PE при рассмотрении следующей выборки. Еще лучше было бы получить графическое представление зависимости вероятности вмешательства PE от переменного среднего значения μ_t . Как построить такой график?

Поскольку границы контрольной карты рассчитываются как границы 99%- и 95%-ного интервалов разброса, мы воспользуемся формулами для определения симметричного интервала разброса при вероятности ошибки $\alpha=0.01$ и 0.05 . Они имеют вид:

$$\bar{x}_{unten}^{oben} = \mu \pm u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad \text{для среднего выборочного } \bar{x}.$$

$$s_{unten} = \sqrt{\frac{\chi^2_{f; \frac{\alpha}{2}}}{f} \cdot \sigma^2} \quad \text{и} \quad s_{oben} = \sqrt{\frac{\chi^2_{f; 1-\frac{\alpha}{2}}}{f} \cdot \sigma^2} \quad \text{при } f=n-1 \text{ для разброса } s$$

Средняя линия $\bar{x}$ -карты представляет собой математическое ожидание $\bar{x}$, т.е. μ . Определение математического ожидания s — это уже более тонкая задача. Было бы ошибкой, основываясь на том, что математическое ожидание выборочной дисперсии s^2 равно σ^2 , полагать, что $E(s) = \sigma$. Вместо этого справедливо следующее:

$$E_s := \sqrt{\frac{2}{n-1}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \cdot \sigma$$

Учитывая все вышесказанное, достаточно просто выполнить контрольные карты, приведенные ниже. В карты внесены 20 значений $\bar{x}$ и s , которые вычислены на основании 20 выборок объема $n=7$, симитированных при помощи генератора нормально распределенных случайных чисел *rnorm*.

Ниже приведен график вероятности вмешательства как функции от сдвига математического ожидания (в обоих направлениях). При заданном значении $\mu=500$ вероятность вмешательства составляет $\alpha=0.01$.

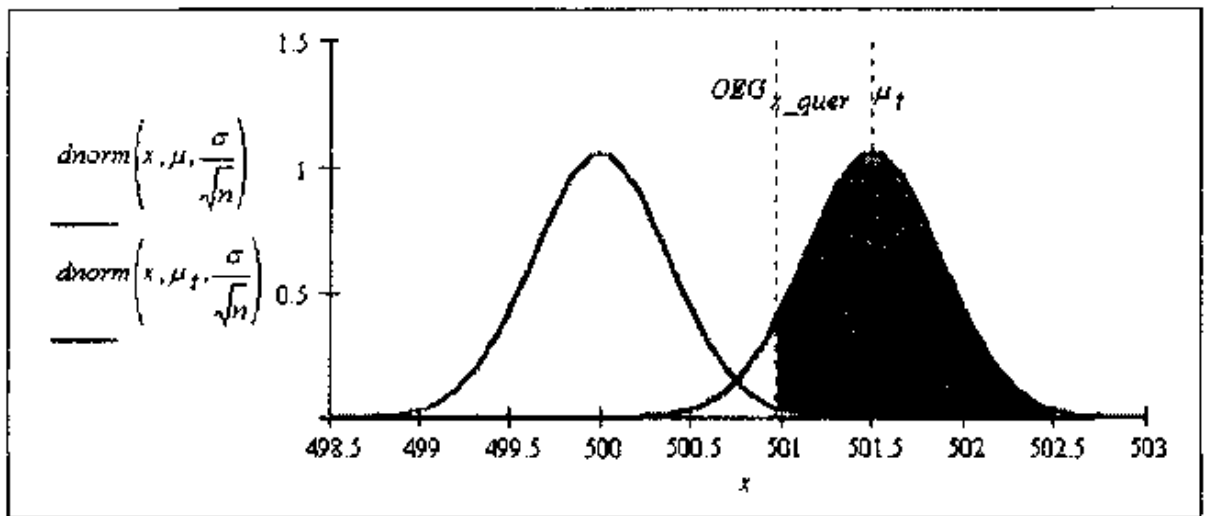


Рис.24 Вероятность вмешательства как функции от сдвига математического ожидания

На рисунке показаны две функции плотности распределения среднего выборочного, в одном случае при заданном значении математического ожидания, в другом — при сдвинутом значении μ_t . Заштрихованная область соответствует вероятности вмешательства. При заданном значении математического ожидания вероятность вмешательства, измеренная справа от верхней границы вмешательства, составляет 0.5%. При перемещении математического ожидания в точку μ_t , она возрастает до 91.8%. Из рисунка видно, каким образом эта вероятность вмешательства может быть вычислена:

$$P_{E(\text{ingriff})} = 1 - P_{A(\text{annahme})} = 1 - (G(OEG_{\bar{x}}, \mu_t, \sigma_{\bar{x}}) - G(UEG_{\bar{x}}, \mu, \sigma_{\bar{x}}))$$

Здесь $G(x, \mu, \sigma)$ — это функция распределения для нормального распределения. В нашем примере $G(UEG, \mu_t, \sigma_{\bar{x}})$ практически равна нулю. Однако эта часть формулы становится самой существенной, если математическое ожидание сдвигается вниз. Таким образом, чтобы эта формула могла использоваться во всех случаях, необходимо учитывать все ее части.

В математическом пакете MathCAD решение будет выглядеть следующем образом:

Объем выборок: $n := 7$

Параметры распределения: $\mu := 500$ $\sigma := 1$

Границы контрольной карты качества (99%-95%-ный интервал разброса) для средних значений:

Математическое ожидание для x_{quer} : $E_{x_{\text{quer}}} := \mu$

$$\begin{aligned}
OEG_{x_quer} &:= qnorm(0.995, \mu, \frac{\sigma}{\sqrt{n}}) & OEG_{x_quer} &= 500.974 \\
OWG_{x_quer} &:= qnorm(0.975, \mu, \frac{\sigma}{\sqrt{n}}) & OWG_{x_quer} &= 500.741 \\
M_{x_quer} &:= E_{x_quer} & M_{x_quer} &= 500 \\
UWG_{x_quer} &:= qnorm(0.025, \mu, \frac{\sigma}{\sqrt{n}}) & UWG_{x_quer} &= 499.259 \\
UEG_{x_quer} &:= qnorm(0.005, \mu, \frac{\sigma}{\sqrt{n}}) & UEG_{x_quer} &= 499.026
\end{aligned}$$

Границы контрольной карты качества (99%-95%-ный интервал разброса):

$$E_s := \sqrt{\frac{2}{n-1}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \cdot \sigma$$

Математическое ожидание для s:

$$\begin{aligned}
OEG_s &:= \sqrt{\frac{qchisq(0.995, n-1)}{n-1}} \cdot \sigma & OEG_s &= 1.758 \\
OWG_s &:= \sqrt{\frac{qchisq(0.975, n-1)}{n-1}} \cdot \sigma & OWG_s &= 1.552 \\
M_s &:= E_s & M_s &= 0.959 \\
UWG_s &:= \sqrt{\frac{qchisq(0.025, n-1)}{n-1}} \cdot \sigma & UWG_s &= 0.454 \\
UEG_s &:= \sqrt{\frac{qchisq(0.005, n-1)}{n-1}} \cdot \sigma & UEG_s &= 0.336
\end{aligned}$$

Имитация 20-ти выборок при помощи генератора случайных чисел с нормальным распределением $rnorm$ (число случайных компонент вектора, μ, σ)

Число выборок: $N := 20$

Промежуток изменения: $i := 0..N - 1$

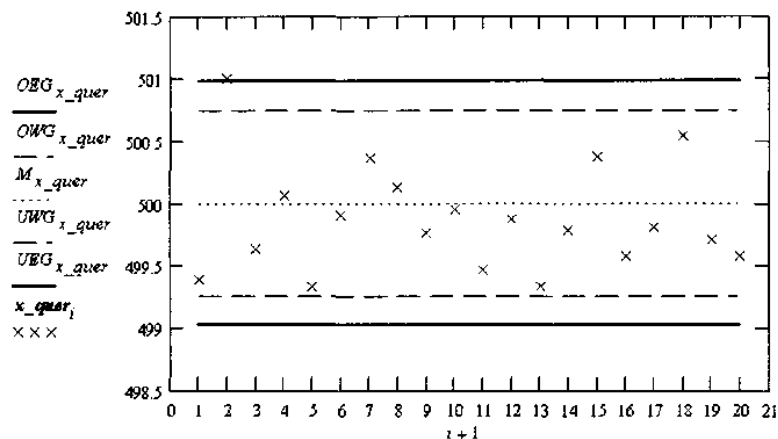
Случайные числа записываем в векторы-столбцы

матрицы X : $x^{<i>} := rnorm(n, \mu, \sigma)$

Среднее выборочное: $x_quer_i = mean(X^{<i>})$

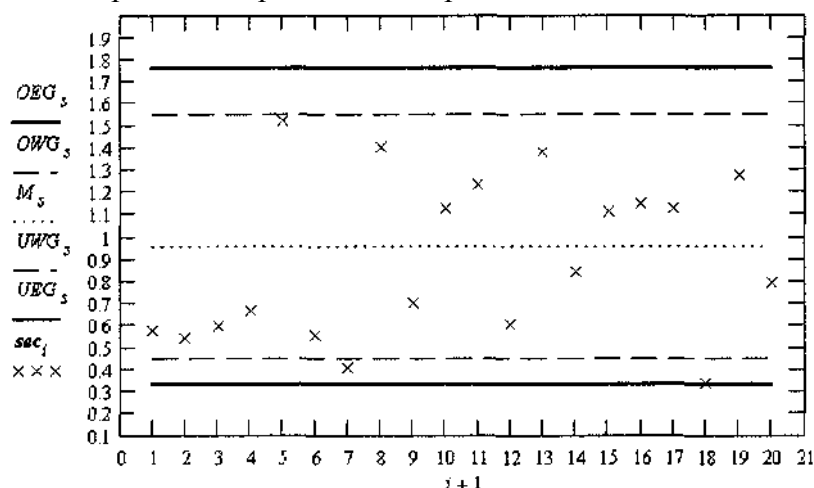
Выборочное среднее квадратичное отклонение: $S_i := \sqrt{\frac{n}{n-1}} stdev(X^{<i>})$

Карта средних значений с границами вмешательства, предупреждающими границами и 20-ю средними выборочными значениями:



Результат: вмешательство при выборке 2!

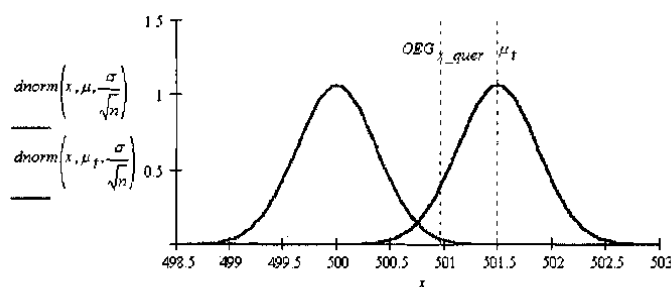
Карта средних квадратичных отклонений с границами вмешательства, предупреждающими границами и 20-ю выборочными средними квадратичными отклонениями:



Результат: предупреждение при выборке 7; вмешательство при выборке 18! Вероятность вмешательства при сдвиге математического ожидания: Смещенное математическое ожидание: $\mu_{it} := 501.5$

Представление этого сдвига математического ожидания посредством функций плотности для мат. ожидания; вероятность вмешательства:

Промежуток изменения для математического ожидания: $x := 498.5, 498.55 \dots 503$



Вычисление вероятности вмешательства P_E :

$$P_E := (1 - \text{pnorm}(\text{OEG}_{x_quer}, \mu_{it}, \frac{\sigma}{\sqrt{n}})) - \text{pnorm}(\text{UEG}_{x_quer}, \mu_{it}, \frac{\sigma}{\sqrt{n}}))$$

Результат: $P_E = 91.816\%$ Сдвиг математического ожидания выявляется при взятии следующей выборки почти с 92%-ной вероятностью.

Представление вероятности вмешательства

как функции математического ожидания:

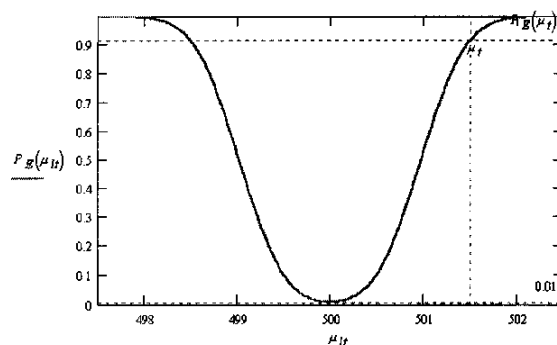
Промежуток изменения для математического ожидания:

$$\mu_{it} := 496.5, 496.52 \dots 503.5$$

Вычисление вероятности вмешательства $P_E(\mu_{it})$:

$$P_E(\mu_{it}) := 1 - (\text{pnorm}(\text{OEG}_{x_quer}, \mu_{it}, \frac{\sigma}{\sqrt{n}}) - \text{pnorm}(\text{UEG}_{x_quer}, \mu_{it}, \frac{\sigma}{\sqrt{n}}))$$

График с выделенным значением при $\mu_{it} = 501.5$:



Практическая работа №8 Построение контрольных карт качества в статистическом методе управления качеством с помощью математического пакета Mathcad и статистического пакета STADIA.

Что бы вы стали делать как специалист по контролю за качеством, если бы рабочий, работающий с расфасовочным устройством из примера 1, отказался вести контрольные карты качества? Он мог бы (хотя и неохотно) вычислять средние значения, но корни при вычислении разброса для него неприемлемы! Его можно либо направить на курсы повышения квалификации, либо перейти при контроле качества к использованию карт исходных значений. Эти карты значительно проще заполнять, т.к. в них указывается вес всех семи упаковок из выборки (исходные значения). Если *все* значения оказываются внутри предупреждающих границ, предполагается, что процесс протекает безукоризненно.

Задание №1

В МП MathCAD вам необходимо спроектировать карту исходных значений, которая позволяла бы работать с тем же объемом выборки ($n=7$). Разработанную карту необходимо изобразить графически с нанесением на нее значений 20-ти выборок, симитированных при помощи генератора случайных чисел. Следует также, как и в предыдущем примере, исследовать поведение вероятности вмешательства.

Задание №2

1. В статистическом пакете STADIA сгенерируйте 20 выборок (Преобразование\Генератор\нормальное распределение с параметрами из задания №1);
2. Изобразите графически полученные данные (статистические методы\контрольные карты)
3. исследуйте полученные данные так же как и в задании №1, исследуйте поведение вероятности вмешательства.

Тест для самоконтроля

1. Выберите правильный ответ. Для контроля отклонения от среднего значения непрерывной переменной необходимо построить:
 - а). R-карта
 - б). X-карта
 - в). S-карта
 - г). C-карта
 - д). U-карта
2. Выберите правильный ответ. Для контроля за степенью изменчивости непрерывной переменной необходимо построить:
 - а). R-карта
 - б). X-карта

- в). S-карта
 - г). C-карта
 - д). U-карта
3. В каких контрольных картах строится график числа дефектов?
 4. Что представляет собой Средняя линия $\bar{x}$ -карты?
 5. Продолжите предложение. Границы контрольных карт качества представляют собой границы...
 6. Вычислите:

$$s_{unten} = \sqrt{\frac{\chi^2_{f; \frac{\alpha}{2}}}{f} \cdot \sigma} \text{ и } s_{oben} = \sqrt{\frac{\chi^2_{f; 1 - \frac{\alpha}{2}}}{f} \cdot \sigma} \text{ при } f=20 \text{ для разброса, } \sigma = 1.$$

ЛИТЕРАТУРА

Основная:

1. Кочетков Е.С., Смерчинская С.О., Соколов В.В. Теория вероятностей и математическая статистика: учебник / Е.С. Кочетков, С.О. Смерчинская, В.В. Соколов. – 2-е изд. – М.: ФОРУМ, 2011. – 240 с. – (Профессиональное образование).
2. Спирина М.С. Теория вероятностей и математическая статистика: учебник для студ. учреждений сред. проф. образования / М.С. Спирина, П.А. Спирин. – М.: Издательский центр «Академия», 2007.

Дополнительная:

1. Гмурман В.Е. Теория вероятностей и математическая статистика / В.Е. Гмурман. – М.: «Высшая школа», 2008.
2. Иванов В.С. Основы математической статистики. - М.: «ФиС», 2006. – 165с.
3. Калинина В.Н. Папкин В.Ф. Математическая статистика. – М.: «Высшая школа», 2005. – 335с.
4. Калинина В.Н., Панкин В.Ф. Математическая статистика / В.Н. Калинина, В.Ф. Панкин. – М.: «Высшая школа», 2009.
5. Макаров Е. Г. Mathcad. Пакет аналитических и численных расчетов для нематематиков: учебный курс. / Е. Г. Макаров. – СПб.: «БХВ-Петербург», 2009. – 384 с.

Учебное текстовое электронное издание

Федосеева Юлия Валерьевна

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Издается полностью в авторской редакции

1,13 Мб

1 электрон. опт. диск

г. Магнитогорск, 2014 год

ФГБОУ ВПО «МГТУ»

Адрес: 455000, Россия, Челябинская область, г. Магнитогорск,
пр. Ленина 38

ФГБОУ ВПО «Магнитогорский государственный
технический университет им. Г.И. Носова»

Многопрофильный колледж

Центр электронных образовательных ресурсов и
дистанционных образовательных технологий

e-mail: ceor_dot@mail.ru