МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«БРЯНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»

Кафедра математики, физики и информатики

Рыжик В.Н.

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания для самостоятельной работы по дисциплине «Высшая математика»

для бакалавров направлений подготовки

35.03.06 Агроинженерия
23.03.02 Наземные транспортно-технологические комплексы
20.03.01 Техносферная безопасность
19.03.04 Технология продукции и организация общественного питания

Брянская область, 2018 г.

УДК 519.21 (07)
ББК 22.171
Р 93

Рыжик, В.Н. Теория вероятностей и математическая статистика: методические указания для самостоятельной работы по дисциплине «Высшая математика» для бакалавров направлений подготовки 35.03.06 Агроинженерия, 23.03.02 Наземные транспортно-технологические комплексы, 20.03.01 Техносферная безопасность, 19.03.04 Технология продукции и организация общественного питания / В.Н. Рыжик. - Брянск: Изд-во Брянский ГАУ, 2018. - 79 с.

Рецензент: доцент кафедры математики, физики и информатики, кандидат технических наук Е.А. Ракул.

Рекомендовано к изданию учебно-методической комиссией факультета Энергетики и природопользования, протокол №6 от 10.04.2018 г.

Оглавление

Введение

Теория вероятностей изучает вероятностные закономерности массовых однородных случайных событий. У нее нет цели, что-либо угадать. Однако, если эксперимент продолжать много раз в идентичных условиях, то можно проследить некоторую закономерность, описываемую определенными законами.

Вокруг каждого из нас летают молекулы воздуха. Некоторые из них обладают высокой, некоторые средней, а некоторые – низкой скоростью. Не имеет смысла угадывать скорость отдельно взятых молекул; но их массовый учёт находит самое широкое применение в теоретических и прикладных физических исследованиях. Обратите внимание, что самолёты «умеют» летать, газовые и паровые котлы обычно не взрываются, а чайники при кипении не скачут по кухне. За многими и многими, казалось бы, обыденными фактами и событиями кроются серьёзные вероятностно-статистические расчёты.

Или пример попроще. Если вы приобретёте лотерейный билет, то вряд ли что-то выиграете и совсем невероятно, что сорвёте крупный куш. Но организатор лотереи даже при случайном розыгрыше тиража *(извлечение пронумерованных шариков и т.п. либо если участники сами угадывают номера)* гарантированно и с высокой точностью знает, сколько билетов выиграют/проиграют, и, понятно, остаётся в прибыли. Лотереи часто называют обманом, однако парадокс состоит в том, что эта гарантия строго обоснована теорией; равно, как и житейская фраза «всё равно ничего не выиграю».

Да, кстати подумайте ещё над одной насущной задачей: многие из нас за жизнь сдают десятки экзаменов, и практически всегда имеет место следующая ситуация: часть вопросов студент знает (либо заготовлены шпоры), а часть вопросов – не знает (либо плавает как мастер спорта). Наступает день «Х»: утро, коридор с 10-15 однокурсниками и дверь, за которой на столе лежит полный комплект билетов. В каком случае вероятнее сдать экзамен – если идти «в первых рядах», «в серединке» или если зайти в аудиторию в числе последних? …Изучаем теорию вероятностей!

Глава 1

Основные понятия теории вероятностей. Испытания и события. Виды случайных событий

Случайным называется событие, если при реализации установленной совокупности условий S оно может либо произойти, либо не произойти. При этом событие будет рассматриваться как результат испытания.

Например, человек стреляет по мишени, которая разделена на четыре части. Тогда выстрел — это испытание, а попадание в конкретную область мишени есть событие. Если появление одного из событий исключает появление других событий в одном и том же испытании, то события называются несовместными. Например, была подброшена монета, при этом появление герба исключает появление надписи. Значит события «выпадение герба» и «выпадение надписи» — совместные. Если в результате испытания появится хотя бы одно из событий, то можно говорить, что несколько событий образуют полную группу. Следовательно, если события, которые образуют полную группу, попарно несовместны, то при испытании появится одно и только одно из этих событий. Например, человек старался выстрелить по цели. Непременно случится одно из событий: попадание или промах. Значит можно сказать, что эти два несовместных события образуют полную группу.

Равновозможными называют события, если есть повод полагать, что ни одно из них не является более возможным, чем другое. Например, при бросании монеты выпадение «герба» и выпадение надписи являются равновозможными событиями. Ведь монета правильной цилиндрической формы изготовлена из однородного материала, а присутствие чеканки не оказывает влияния на выпадение той или иной стороны монеты.

§1. Вероятность события

Опр. Вероятность события А — это отношение числа благоприятствующих этому событию исходов к общему числу всех равновозможных несовместных элементарных исходов, которые образуют полную группу:

$P(A) = m / n$, - формула классической вероятности

где m — число элементарных исходов, которые благоприятствуют А;

n — число всех возможных элементарных исходов испытания.

При этом предполагается, что элементарные исходы несовместны, равновозможны и образуют полную группу. Следовательно, можно записать следующие три свойства.

1. Вероятность достоверного события равна единице. Следовательно, если событие достоверно, то каждый элементарный исход испытания благоприятствует событию, тогда $m = n$, и $P(A) = m / n = n / n = 1$.

2. Вероятность невозможного события равна нулю. Следовательно, если событие невозможно, то ни один из элементарных исходов испытания не благоприятствует событию, тогда $m = 0$, и $P(A) = m / n = 0 / n = 0$.

3. Вероятность случайного события есть положительное число, заключенное между нулем и единицей. Следовательно, слу□ чайному событию благоприятствует лишь часть из общего числа элементарных исходов испытания, тогда $0 < m < n$, стало быть, $0 < m / n < l$, и $0 < P(A) < 1$ и $0 \leq P(A) \leq 1$.

§2. Относительная частота. Устойчивость относительной частоты. Статистическая вероятность

Относительная частота события — это отношение числа испытаний, в которых событие появилось, к общему числу фактически произведенных испытаний:

$W(A) = m / n$, где m — число появлений события; n — общее число испытаний.

Следует отметить, что для определения вероятности необязательно, чтобы испытания производились в действительности. Для определения относительной частоты необходимо, чтобы испытания были произведены фактически. Значит, можно сказать, что вероятность вычисляют до опыта, а относительную частоту — после опыта.

Например, сделали 35 выстрелов по цели, при этом было зарегистрировано 25 попаданий. Значит, относительная частота поражения цели равна: $W(A) = 25/35$. В различных опытах относительная частота изменяется мало (тем меньше, чем больше произведено испытаний), колеблясь около некоторого постоянного числа. Этим постоянным числом является вероятность появления события. За статистическую вероятность события принимают относительную частоту или число, близкое к ней. В том случае, если событие достоверно, то $m = n$, а относительная частота равна: $m / n = n / n = 1$. Значит, статистическая вероятность достоверного события равна единице.

Если событие невозможно, то $m = 0$, а относительная частота равна: $0 / n = 0$. Следовательно, статистическая вероятность невозможного события равняется нулю.

Для любого события $0 \leq m \leq n$ относительная частота равна: $0 \leq m / n \leq 1$. Другими словами, статистическая вероятность любого события лежит в пределах между нулем и единицей. Для наличия статистической вероятности события A необходимо:

1) чтобы была возможность производить ограниченное число испытаний, в каждом из которых событие A наступает или не наступает;

2) чтобы существовала устойчивость относительных частот появления события. A в различных сериях довольно большого числа испытаний.

§3. Геометрическая вероятность

Недостаток классического определения вероятности состоит в том, что оно неприменимо к испытаниям с бесконечным числом исходов. Для того чтобы его преодолеть, вводят геометрические вероятности — вероятности попадания точки в область (например, отрезок, часть плоскости). Допустим, отрезок l есть часть отрезка L, на котором наудачу поставлена точка. Это значит, что могут выполняться следующие предположения: точка может оказаться в любой точке отрезка L, вероятность попадания точки на отрезок l пропорциональна длине этого отрезка и не зависит от его расположения относительно отрезка L. Тогда вероятность попадания точки на отрезок l равна:

P = Длина l / Длина L.

Аналогично

P = Площадь k / Площадь K

Вероятность попадания брошенной точки на фигуру k пропорциональна площади этой фигуры и не зависит ни от ее расположения относительно K, ни от формы k.

§4. Элементы комбинаторики

Перестановки. Сочетания. Размещения

Опр. Перестановками из n элементов, называют такие соединения, которые состоят из всех элементов n. Одно соединение отличается от другого только порядком расположения этих элементов.
Количество всех возможных перестановок можно вычислить по формуле:

$$\underline{P_n = n!}$$

Задача.

Сколько существует способов выстроить очередь к банкомату из 5 человек?

Решение

$P_n = n!$-формула, n=5, $P_5 = 5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$

Ответ: всего существует 120 способов.

Задача.

Сколько четырехзначных чисел можно составить из карточек с цифрами: 0; 3; 5; 9?

Решение

Всего 4 карточки с цифрами, они все участвуют в перестановке. Применяем формулу:

$P_n = n!,$ n=4, $P_4 = 4! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot = 24$

Ответ: всего существует 24 способа.

Опр. Сочетаниями из данных n элементов называются такие соединения, состоящие из m элементов, взятых из n элементов, каждое из которых, отличается друг от друга, хотя бы одним элементом, в каждом таком соединении порядок элементов неважен.

Обозначаются сочетания: C_n^m и вычисляются по формуле:

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

Задача.

В ящике находится 15 деталей. Сколькими способами можно взять из ящика 4 детали?

Решение

Всего 15 деталей: n=15

Взяли из них 4 детали, при чем неважно, какая первая, вторая и т.д. деталь:

m=4

$$C_n^m = \frac{n!}{(n-m)! \cdot m!}$$

$$C_{15}^4 = \frac{15!}{(15-4)!4!} = \frac{15!}{11!4!} = \frac{12 \cdot 13 \cdot 14 \cdot 15}{1 \cdot 2 \cdot 3 \cdot 4} = 1365$$

Ответ: всего 1365 способов существует взять 4 детали из 15.

Опр. Размещениями называют такие соединения, состоящие из m элементов, взятых из данных n элементов, которые отличаются друг от друга, как составом элементов, так и их порядком. Обозначаются размещения: A_n^m и вычисляются по формуле:

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

Задача.

Сколько существует способов составления 2-значных чисел из цифр: 1; 3; 5; 9?

Решение

Всего 4 цифры n=4

Из них составляют двузначные числа, т.е. m=2. Порядок цифры имеет значение (19 или 91)

$$A_n^m = \frac{n!}{(n-m)!} = \frac{4!}{(4-2)!} = \frac{4!}{2!} = 3 \cdot 4 = 12$$

Задача.

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

Решение

В данном случае не годится подсчёт C_{23}^2 , поскольку общее количество сочетаний включает в себя и разнополые пары.

Условие «выбрать двух человек одного пола» подразумевает, что необходимо выбрать двух юношей или двух девушек, и уже сама словесная формулировка указывает на верный путь решения:

$C_{10}^2 = \frac{10!}{8! \cdot 2!} = \frac{8! \cdot 9 \cdot 10}{8! \cdot 2!} = \frac{9 \cdot 10}{2} = 45$ - способами можно выбрать 2 юношей;

$C_{13}^2 = \frac{13!}{11! \cdot 2!} = \frac{12 \cdot 13}{2} = 6 \cdot 13 = 78$ - способами можно выбрать 2 девушек.

Таким образом, двух человек одного пола (без разницы – юношей или девушек) можно выбрать: $C_{10}^2 + C_{13}^2 = 45 + 78 = 123$ способами (знак «+» следует понимать, как союз «или»)

Ответ: 123

Рассмотрим ту же студенческую группу, которая пошла на танцы. Сколькими способами можно составить пару из юноши и девушки?

$$C_{10}^1 = 10$$

способами можно выбрать 1 юношу;

$$C_{13}^1 = 13$$

способами можно выбрать 1 девушку.

Таким образом, одного юношу **и** одну девушку можно выбрать: $C_{10}^1 \cdot C_{13}^1 = 10 \cdot 13 = 130$ способами (знак «•» следует понимать, как союз «и»).

Задача.

Сколько существует трёхзначных чисел, которые делятся на 5?

Решение

Для наглядности обозначим данное число тремя звёздочками: *******

Комбинации будем считать по разрядам – *слева направо*:

В *разряд сотен* можно записать любую из $C_9^1 = 9$ цифр (1, 2, 3, 4, 5, 6, 7, 8 или 9). Ноль не годится, так как в этом случае число перестаёт быть трёхзначным.

А вот в *разряд десятков* («посерединке») можно выбрать любую из 10 цифр: $C_{10}^1 = 10$.

По условию, число должно делиться на 5. Число делится на 5, если оно заканчивается на 5 либо на 0. Таким образом, в младшем разряде нас устраивают 2 цифры.

Итого, существует:

$$C_9^1 \cdot C_{10}^1 \cdot 2 = 9 \cdot 10 \cdot 2 = 180$$

трёхзначных чисел, которые делятся на 5.

При этом произведение $C_9^1 \cdot C_{10}^1 \cdot 2$ расшифровывается так: «9 способами можно выбрать цифру в *разряд сотен,* и 10 способами выбрать цифру в *разряд десятков,* и 2 способами в *разряд единиц*»

Или ещё проще: «каждая из 9 цифр в *разряде сотен* комбинируется с каждой из 10 цифр *разряда десятков* и с каждой из двух цифр в *разряде единиц*».

Ответ: 180

§5. Перестановки, сочетания, размещения с повторениями

Перестановки с повторениями

В перестановках с повторениями, как и в «обычных» перестановках, участвует сразу всё множество объектов, но есть одно но: в данном множестве один или большее количество элементов (объектов) повторяются. Встречайте очередной стандарт:

Задача.

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

Решение

В том случае, если бы все буквы были различны, то следовало бы применить тривиальную формулу P_n, однако совершенно понятно, что для предложенного набора карточек некоторые манипуляции будут срабатывать «вхолостую», так, например, если поменять местами любые две карточки с буквами «К» в любом слове, то получится то же самое слово. Причём, физически карточки могут сильно отличаться: одна быть круглой с напечатанной буквой «К», другая – квадратной с нарисованной буквой «К». Но по смыслу задачи даже такие карточки считаются одинаковыми, поскольку в условии спрашивается о буквосочетаниях.

Всё предельно просто – всего: 11 карточек, среди которых буква:

К – повторяется 3 раза;
О – повторяется 3 раза;
Л – повторяется 2 раза;
Ь – повторяется 1 раз;
Ч – повторяется 1 раз;
И – повторяется 1 раз.

Проверка: 3 + 3 + 2 + 1 + 1 + 1 = 11, что и требовалось проверить.

По формуле

$$P_{11(повт)} = \frac{n!}{n_1! \cdot n_2! \cdot n_3! \cdot \ldots \cdot n_k!} = \frac{11!}{3! \cdot 3! \cdot 2! \cdot 1! \cdot 1! \cdot 1!} = \frac{39916800}{6 \cdot 6 \cdot 2} = 554400$$

различных буквосочетаний можно получить. Больше полумиллиона!

На практике вполне допустимо не записывать общую формулу и, кроме того, опускать единичные факториалы:

$$P_{11(повт)} = \frac{11!}{3! \cdot 3! \cdot 2!} = \frac{39916800}{6 \cdot 6 \cdot 2} = 554400$$

Ответ: 554400

Другой типовой пример перестановок с повторениями встречается в задаче о расстановке шахматных фигур.

Для самостоятельного решения менее шаблонное задание:

Задача.

Алексей занимается спортом, причём 4 дня в неделю – лёгкой атлетикой, 2 дня – силовыми упражнениями и 1 день отдыхает. Сколькими способами он может составить себе расписание занятий на неделю?

Формула $P_7 = 7!$ здесь не годится, поскольку учитывает совпадающие перестановки (например, когда меняются местами силовые упражнения в среду с силовыми упражнениями в четверг). И опять – по факту те же 2 силовые тренировки могут сильно отличаться друг от друга, но по контексту задачи (с точки зрения расписания) они считаются одинаковыми элементами.

Сочетания с повторениями

Характерная особенность этого вида комбинаций состоит в том, что выборка проводится из нескольких групп, каждая из которых состоит из одинаковых объектов.

Задача

В студенческой столовой продают сосиски в тесте, ватрушки и пончики. Сколькими способами можно приобрести пять пирожков?

Решение

Физические характеристики пирожков по смыслу задачи не существенны, и хот-доги / ватрушки / пончики в своих группах считаются одинаковыми.

Что может быть в выборке? Прежде всего, следует отметить, что в выборке обязательно будут одинаковые пирожки (т.к. выбираем 5 штук, а на выбор предложено 3 вида). Варианты тут на любой вкус: 5 хот-догов, 5 ватрушек, 5 пончиков, 3 хот-дога + 2 ватрушки, 1 хот-дог + 2 + ватрушки + 2 пончика и т.д.

Как и при «обычных» сочетаниях, порядок выбора и размещение пирожков в выборке не имеет значения – просто выбрали 5 штук и всё.

Используем формулу:

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

количества сочетаний с повторениями:

$$C_{3(\text{повт})}^5 = C_{3+5-1}^5 = C_7^5 = \frac{7!}{2!\cdot 5!} = \frac{6\cdot 7}{2!} = 21$$

способом можно приобрести 5 пирожков.

Ответ: 21

Какой вывод можно сделать из многих комбинаторных задач?

Порой, самое трудное – это разобраться в условии.

Аналогичный пример для самостоятельного решения:

Задача.

В кошельке находится достаточно большое количество рублей, 2-, 5- и 10-рублёвых монет. Сколькими способами можно извлечь три монеты из кошелька?

В целях самоконтроля ответьте на пару простых вопросов:

1) Могут ли в выборке все монеты быть разными?

2) Назовите самую «дешевую» и самую «дорогую» комбинацию монет.

Размещения с повторениями

Из множества, состоящего из n элементов, выбирается m элементов, при этом важен порядок элементов в каждой выборке. И всё бы было ничего, но довольно неожиданный прикол заключается в том, что любой объект исходного множества мы можем выбирать сколько угодно раз. Образно говоря, от «множества не убудет».

Когда так бывает? Типовым примером является кодовый замок с несколькими дисками, но по причине развития технологий актуальнее рассмотреть его цифрового потомка:

Задача.

Сколько существует четырёхзначных пин-кодов?

Решение

для решения задачи достаточно знаний правил комбинаторики: $C_{10}^1 = 10$ способами можно выбрать первую цифру пин-кода и $C_{10}^1 = 10$ способами – вторую цифру пин-кода, и столькими же способами – третью, **и** столькими же – четвёртую. Таким образом, по правилу умножения комбинаций, четырёхзначный пин-код можно составить: $C_{10}^1 \cdot C_{10}^1 \cdot C_{10}^1 \cdot C_{10}^1 = 10000$ способами.

А теперь с помощью формулы. По условию нам предложен набор из $n = 10$ цифр, из которого выбираются $m = 4$ цифры и располагаются <u>в определенном порядке</u>, при этом цифры в выборке могут повторяться *(т.е. любой цифрой исходного набора можно пользоваться произвольное количество раз)*. По формуле $A_{n(\text{повт})} = n^m$ количества размещений с повторениями:

$$A^4_{10(\text{повт})} = 10^4 = 10000$$

Ответ: 10000

Согласно государственному стандарту, автомобильный номерной знак состоит из 3 цифр и 3 букв. При этом недопустим номер с тремя нулями, а буквы выбираются из набора А, В, Е, К, М, Н, О, Р, С, Т, У, Х *(используются только те буквы кириллицы, написание которых совпадает с латинскими буквами)*.

Сколько различных номерных знаков можно составить для региона?

§6. Задачи на классическое определение вероятности

Задача

В урне находится 15 белых, 5 красных и 10 чёрных шаров. Наугад извлекается 1 шар, найти вероятность того, что он будет: а) белым, б) красным, в) чёрным.

Решение

Всего в урне: 15 + 5 + 10 = 30 шаров, и, очевидно, справедливы следующие факты:

– извлечение любого шара одинаково возможно *(равновозможность исходов)*, при этом исходы *элементарны* и образуют *полную группу событий (т.е. в результате испытания обязательно будет извлечён какой-то один из 30-ти шаров)*.

Таким образом, общее число исходов: $n = 30$

Рассмотрим событие: A – из урны будет извлечён белый шар. Данному событию благоприятствуют m=15 элементарных исходов, поэтому по классическому определению:

$P(A) = \frac{m}{n} = \frac{15}{30} = \frac{1}{2}$ – вероятность того, то из урны будет извлечён белый шар.

С другими пунктами аналогично, рассмотрим следующие события:

B – из урны будет извлечён красный шар; C – из урны будет извлечён чёрный шар.

Событию B благоприятствует 5 элементарных исходов, а событию C – 10 элементарных исходов. Таким образом, соответствующие вероятности:

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

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

Типичная проверка многих задач осуществляется с помощью теоремы о сумме вероятностей событий, образующих полную группу. В нашем случае события A,B,C образуют полную группу, а значит, сумма соответствующих вероятностей должна обязательно равняться единице: $P(A)+P(B)+P(C)=1$.

Проверим, так ли это:

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

в чём и хотелось убедиться.

Ответ: а) $\frac{1}{2}$ b) $\frac{1}{6}$ в) $\frac{1}{3}$

Задача

Набирая номер телефона, абонент забыл две последние цифры, но помнит, что одна из них – ноль, а другая – нечётная. Найти вероятность того, что он наберёт правильный номер.

Примечание*: ноль – это чётное число (делится на 2 без остатка)*

Решение

Сначала найдём общее количество исходов. По условию, абонент помнит, что одна из цифр – ноль, а другая цифра – нечётная. Здесь воспользоваться *методом прямого перечисления исходов*. То есть, при оформлении решения просто записываем все комбинации:

01, 03, 05, 07, 09

10, 30, 50, 70, 90

И подсчитываем их – всего: 10 исходов.

Благоприятствующий исход один: верный номер.

По классическому определению:

$p=\frac{10}{100}=\frac{1}{10}=0,1$ – вероятность того, что абонент наберёт правильный номер

Ответ: 0,1

Задача

Найти вероятность того, что при бросании двух игральных костей в сумме выпадет:

а) пять очков;

б) не более четырёх очков;

в) от 3-х до 9 очков включительно.

Решение

Найдём общее количество исходов: $C_6^1 = 6$ способами может выпасть грань 1-го кубика и $C_6^1 = 6$ способами может выпасть грань 2-го кубика; по правилу умножения комбинаций, всего: $C_6^1 \cdot C_6^1 = 6 \cdot 6 = 36$ возможных комбинаций. Иными словами, каждая грань 1-го кубика может составить <u>упорядоченную</u> пару с каждой гранью 2-го кубика. Условимся записывать такую пару в виде (a, b), где a – цифра, выпавшая на 1-м кубике, b – цифра, выпавшая на 2-м кубике. Например:

$(3, 5)$ – на первом кубике выпало 3 очка, на втором – 5 очков, сумма очков: 3 + 5 = 8;

$(6, 1)$ – на первом кубике выпало 6 очков, на втором – 1 очко, сумма очков: 6 + 1 = 7;

$(2, 2)$ – на обеих костях выпало 2 очка, сумма: 2 + 2 = 4.

Очевидно, что наименьшую сумму даёт пара (1;1) , а наибольшую – две «шестёрки».

а) Рассмотрим событие: A – при бросании двух игральных костей выпадет 5 очков. Запишем и подсчитаем количество исходов, которые благоприятствуют данному событию: $(1, 4); (4, 1); (2, 3); (3, 2)$.

Итого: 4 благоприятствующих исхода. По классическому определению:

$P(A) = \frac{4}{36} = \frac{1}{9}$ – искомая вероятность.

б) Рассмотрим событие: B – выпадет не более 4-х очков. То есть, либо 2, либо 3, либо 4 очка. Снова перечисляем и подсчитываем благоприятствующие комбинации, слева я буду записывать суммарное количество очков, а после двоеточия – подходящие пары:

$2\ \textit{очка}: \quad (1, 1)$

$3\ \textit{очка}: \quad (1, 2); (2, 1)$

$4\ \textit{очка}: \quad (1, 3); (3, 1); (2, 2)$

Итого: 6 благоприятствующих комбинаций. Таким образом:

$P(B) = \frac{6}{36} = \frac{1}{6}$ – вероятность того, что выпадет не более 4-х очков.

в) Рассмотрим событие: С – выпадет от 3-х до 9 очков включительно. Здесь можно пойти прямой дорогой, но… что-то не хочется. Да, некоторые пары уже перечислены в предыдущих пунктах, но работы все равно предстоит многовато.

Как лучше поступить? В подобных случаях рациональным оказывается окольный путь. Рассмотрим противоположное событие $\overline{C}$ – выпадет 2 или 10 или 11 или 12 очков.

В чём смысл? Противоположному событию благоприятствует значительно меньшее количество пар:

$2\ \textit{очка}:\quad (1,1)$

$10\ \textit{очков}:\quad (4,6);\ (6,4);\ (5,5)$

$11\ \textit{очков}:\quad (5,6);\ (6,5)$

$12\ \textit{очков}:\quad (6,6)$

Итого: 7 благоприятствующих исходов.

По классическому определению:

$P(\overline{C})=\frac{7}{36}$ – вероятность того, что выпадет меньше трёх или больше 9-ти очков.

Далее пользуемся тем, что сумма вероятностей противоположных событий равна единице:

$$P(C)+P(\overline{C})=1\quad\Rightarrow\quad P(C)=1-P(\overline{C})=1-\frac{7}{36}=\frac{29}{36}$$

– вероятность того, что выпадет от 3-х до 9 очков включительно.

Особо щепетильные люди могут перечислить все 29 пар, выполнив тем самым проверку.

Ответ: а) $\frac{1}{9}$, b) $\frac{1}{6}$,в) $\frac{29}{36}$

Задача

В лифт 20-этажного дома на первом этаже зашли 3 человека. И поехали. Найти вероятность того, что:

а) они выйдут на разных этажах
б) двое выйдут на одном этаже;
в) все выйдут на одном этаже.

Следует отметить, что случайность здесь имеет место быть лишь с точки зрения стороннего наблюдателя *(т.к. человек обычно едет на вполне определённый этаж)*.

Решение

Вычислим общее количество исходов: $C_{19}^{1}=19$ способами может выйти из лифта 1-й пассажир **и** $C_{19}^{1}=19$ способами – 2-й пассажир **и** $C_{19}^{1}=19$ способами – третий пассажир. По правилу умножения комбинаций: $C_{19}^1 \cdot C_{19}^1 \cdot C_{19}^1 = 19 \cdot 19 \cdot 19 = 6859$ возможных исходов.

То есть, каждый этаж выхода 1-го человека может комбинироваться с каждым этажом выхода 2-го человека и с каждым этажом выхода 3-го человека.

Второй способ основан на размещениях с повторениями: $A_{19(\text{повт})}^3 = 19^3$ – кому как понятнее.

а) Рассмотрим событие: A – пассажиры выйдут на разных этажах. Вычислим количество благоприятствующих исходов:

$A_{19}^3 = 17 \cdot 18 \cdot \cdot 195814$ способами могут выйти 3 пассажира на разных этажах.

Рассуждения по формуле $C_{19}^3 \cdot P_3$ проведите самостоятельно.

По классическому определению:

$$P(A) = \frac{5814}{6859} = \frac{306}{361}$$

в) Рассмотрим событие: B – пассажиры выйдут на одном этаже. Данному событию благоприятствуют $C_{19}^1 = 19$ исходов и по классическому определению, соответствующая вероятность:

$$P(B) = \frac{19}{6859} = \frac{1}{361}.$$

Заходим с чёрного хода:

б) Рассмотрим событие:С – два человека выйдут на одном этаже .

События А,В,С образуют полную группу

$$P(A) + P(B) + P(C) = 1.$$

В результате, искомая вероятность:

$$P(C) = 1 - P(A) - P(B) = 1 - \frac{306}{361} - \frac{1}{361} = \frac{54}{361}$$

Использовали теорему сложения вероятностей.

Ответ: а) $\frac{306}{361} \approx 0{,}8476$, б) $\frac{54}{361} \approx 0{,}1496$, в) $\frac{1}{361} \approx 0{,}0028$

Глава 2

Основные теоремы теории вероятностей

§1. Теорема сложения вероятностей

Опр. События называются несовместными, если появление одного из них исключает появление другого.

Событие, включающее появление события А, или события В, или обоих этих событий, называется суммой А + В двух событий А и В. Если два события А и В являются несовместными, то А + В есть событие, включающее появление одного из этих событий Событие, заключающееся в появлении хотя бы одного из событий, называется суммой нескольких событий.

Теорема. Вероятность появления одного из двух несовместных событий, неважно какого, равняется сумме вероятностей этих событий:

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

Опр. События называются совместными, если появление одного из них, не исключает появление другого.

Для таких событий справедлива теорема сложения вероятностей:

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

Задача

Рассмотрим событие $B_{5,6}$ – в результате броска игральной кости выпадет не менее пяти очков. Данное событие состоит из двух несовместных исходов:

$$B_{5,6} = B_5 + B_6$$

выпадет 5 или *6 очков*.

По теореме сложения вероятностей несовместных событий: $P(B_5 + B_6) = P(B_5) + P(B_6) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

– вероятность того, что в результате броска игральной кости выпадет не менее пяти очков.

Задача

Рассмотрим событие $B_{1-4} = B_1 + B_2 + B_3 + B_4$, состоящее в том, что выпадет не более 4-х очков и найдем его вероятность.

По теореме сложения вероятностей несовместных событий:

$$P(B_1+B_2+B_3+B_4)=P(B_1)+P(B_2)+P(B_3)+P(B_4)=\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{2}{3}$$

По той же теореме, вероятность того, что выпадет нечётное число очков:

$$P(B_1+B_3+B_5)=P(B_1)+P(B_3)+P(B_5)=\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}$$

§2. Зависимые и независимые события. Теоремы умножения вероятностей

Опр. События являются независимыми, если вероятность наступления любого из них не зависит от появления(не появления) остальных событий рассматриваемого множества (во всех возможных комбинациях).

Теорема умножения вероятностей независимых событий

Вероятность совместного появления независимых событий А и В равна произведению вероятностей этих событий:

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

Опр. События называются зависимым*и,* если вероятность появления одного из них, зависит от того, произошло другое событие или нет.

Вероятность события В , вычисленная в предположении того, что событие А уже произошло, называется ***условной вероятностью*** наступления события В и обозначается через $P(A)_B$. При этом события А и В называют ***зависимыми событиями*** *(хотя, строго говоря, зависимо только одно из них)*.

Теорема умножения вероятностей зависимых событий

Вероятность совместного появления двух зависимых событий равна произведению вероятности одного из них на условную вероятность другого, вычисленную в предположении, что первое событие уже произошло: $P(AB)=P(A)\cdot P_A(B)$

Опр. Два события, которые единственно возможны и образуют полную группу, называются противоположными. А- одно из двух противоположных событий, а другое $\overline{A}$.

Так, например, при стрельбе пор мишени попадание в мишени и промах, есть два противоположных события

Задача

В урне 4 белых и 7 черных шаров. Из урны наудачу один за другим извлекают два шара, не возвращая их обратно. Найти вероятность того, что:

а) оба шара будут белыми;
б) оба шара будут чёрными;
в) сначала будет извлечён белый шар, а затем – чёрный.

Решение

Всего в урне: 4 + 7 = 11 шаров.

а) Рассмотрим события A – первый шар будет белым,B – второй шар будет белым и найдём вероятность события AB , состоящего в том, что 1-ый шар будет белым и 2-ой белым.

По классическому определению вероятности:

$$P(A) = \frac{4}{11}.$$

Предположим, что белый шар извлечён, тогда в урне останется 10 шаров, среди которых 3 белых, поэтому:

$P_A(B) = \frac{3}{10}$ – вероятность извлечения белого шара во 2-м испытании при условии, что до этого был извлечён белый шар.

По теореме умножения вероятностей зависимых событий:

$$P(AB) = P(A) \cdot P_A(B) = \frac{4}{11} \cdot \frac{3}{10} = \frac{6}{55}$$

– вероятность того, что оба шара будут белыми.

б) Найдём вероятность события $\overline{A}\overline{B}$, состоящего в том, что 1-ый шар будет чёрным и 2-ой чёрным

По классическому определению:

$$P(\overline{A}) = \frac{7}{11}$$

– вероятность того, что в 1-м испытании будет извлечён чёрный шар.

Пусть извлечён чёрный шар, тогда в урне останется 10 шаров, среди которых 6 чёрных, следовательно:

$$P_{\overline{A}}(\overline{B}) = \frac{6}{10} = \frac{3}{5}$$

– вероятность того, что во 2-м испытании будет извлечён чёрный шар при условии, что до этого был извлечен чёрный шар.

По теореме умножения вероятностей зависимых событий:

$$P(\overline{A}\overline{B}) = P(\overline{A}) \cdot P_{\overline{A}}(\overline{B}) = \frac{7}{11} \cdot \frac{3}{5} = \frac{21}{55}$$

– вероятность того, что оба шара будут чёрными.

в) Найдём вероятность события $A\overline{B}$ (сначала будет извлечён белый шар и затем чёрный)

После извлечения белого шара, в урне останется 10 шаров, среди которых 3 белых и 7 чёрных, таким образом:

$$P_A(\overline{B}) = \frac{7}{10}$$

– вероятность того, что во 2-м испытании будет извлечён чёрный шар при условии, что до этого был извлечен белый шар.

По теореме умножения вероятностей зависимых событий:

$$P(A\overline{B}) = P(A) \cdot P_A(\overline{B}) = \frac{4}{11} \cdot \frac{7}{10} = \frac{14}{55}$$

Ответ:а) $\frac{6}{55}$ б) $\frac{21}{55}$ в) $\frac{14}{55}$

Задача

В первой урне содержится 12 шаров, из них 7 белых, во второй – 6 шаров, из них 3 белых. Из первой урны во вторую наудачу перекладывают один шар, а затем из второй урны наудачу извлекают один шар. Найти вероятность того, что он окажется белым.

Решение

По условию, из первой урны во вторую наудачу перекладывают один шар, и, очевидно, он может быть как белым, так и чёрным. В этой связи необходимо рассмотреть 2 несовместные *гипотезы*:

B_1 – из 1-й урны во 2-ую будет переложен белый шар; B_2 – из 1-й урны во 2-ую будет переложен чёрный шар.

Обозначим через А зависимое событие – из 2-й урны будет извлечён белый шар.

Несовместные исходы удобно расписать по пунктам:

По классическому определению: $P(B_1)\frac{7}{12}$ – вероятность того, что из 1-й урны во вторую будет переложен белый шар. Пусть гипотеза B_1 осуществилась, тогда во второй урне стало 7 шаров, среди которых теперь 4 белых шара.

Таким образом: $P_{B_1}(A)=\frac{4}{7}$ – вероятность того, что из второй урны будет извлечен белый шар при условии, что туда был переложен белый шар.

По теореме умножения вероятностей зависимых событий: $P(B_1A)=P(B_1)\cdot P_{B_1}(A)=\frac{7}{12}\cdot\frac{4}{7}=\frac{1}{3}$ – вероятность того, что во 2-ую урну будет переложен белый шар **и** после этого из 2-ой урны будет извлечён белый шар.

По классическому определению: $P(B_2)=\frac{5}{12}$ – вероятность того, что из 1-й урны во вторую будет переложен чёрный шар. Пусть гипотеза B_2 осуществилась, тогда во второй урне стало 7 шаров, среди которых по-прежнему 3 белых.

Таким образом: $P_{B_2}(A)=\frac{3}{7}$ – вероятность того, что из второй урны будет извлечен белый шар при условии, что туда был переложен чёрный шар.

По теореме умножения вероятностей зависимых событий: $P(B_2A)=P(B_2)\cdot P_{B_2}(A)=\frac{5}{12}\cdot\frac{3}{7}=\frac{5}{28}$ – вероятность того, что из 1-й урны во 2-ую будет переложен чёрный шар **и** после этого из 2-ой урны будет извлечён белый шар.

По теореме сложения вероятностей несовместных событий $P(A)=P(B_1A)+P(B_2A)=\frac{1}{3}+\frac{5}{28}=\frac{43}{84}$

– вероятность того, что из 2-й урны будет извлечён белый шар.

Ответ: $\frac{43}{84}\approx 0{,}51$

§3. Вероятность появления хотя бы одного события

Теорема. Вероятность появления хотя бы одного из событий $A_1, A_2, \ldots A_n$, которые независимы в совокупности, равняется разности между единицей и произведением вероятностей противоположных событий.

P(A) = 1— $g_1 \cdot g_2 \ldots g_n$, где g- вероятность противоположного события.

Формула полной вероятности. Формула Байеса

Теорема. Пусть несовместные события $B_1, B_2 \ldots B_n$ образуют полную группу. Тогда вероятность события А, которое может наступить только при условии появления одного из этих несовместных событий, равняется сумме произведений вероятностей каждого из этих событий на соответствующую условную вероятность события А:

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + P(B_3) \cdot P_{B_3}(A) + \ldots + P(B_n) \cdot P_{B_n}(A)$$

Задача

Имеются три одинаковые урны. В первой урне находятся 4 белых и 7 черных шаров, во второй – только белые и в третьей – только черные шары. Наудачу выбирается одна урна и из неё наугад извлекается шар. Какова вероятность того, что этот шар чёрный?

Решение

Рассмотрим событие A – из наугад выбранной урны будет извлечён чёрный шар. Данное событие может произойти в результате осуществления одной из следующих гипотез:

B_1 – будет выбрана 1-ая урна;

B_2 – будет выбрана 2-ая урна;

B_3 – будет выбрана 3-я урна.

Так как урна выбирается наугад, то выбор любой из трёх урн равновозможен, следовательно:

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$$

Обратите внимание, что перечисленные гипотезы образуют полную группу событий, то есть по условию чёрный шар может появиться только из этих урн, а например, не прилететь с бильярдного стола. Проведём простую промежуточную проверку:

$$P(B_1) + P(B_2) + P(B_3) = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$$

В первой урне 4 белых + 7 черных = 11 шаров,

$$P_{B_1}(A) = \frac{7}{11}$$

– вероятность извлечения чёрного шара при условии, что будет выбрана 1-ая урна.

Во второй урне только белые шары, поэтому в случае её выбора появления чёрного шара становится *невозможным*:

$$P_{B_2}(A) = 0.$$

И, наконец, в третьей урне одни чёрные шары, а значит, соответствующая условная вероятность извлечения чёрного шара составит

$$P_{B_3}(A) = 1$$

(событие достоверно).

По формуле полной вероятности:

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + P(B_3) \cdot P_{B_3}(A) =$$

$$= \frac{1}{3} \cdot \frac{7}{11} + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 = \frac{7}{33} + 0 + \frac{1}{3} = \frac{18}{33} = \frac{6}{11}$$

– вероятность того, что из наугад выбранной урны будет извлечен чёрный шар.

Ответ: $\frac{6}{11}$

Формула Байеса

Пусть событие А наступило в результате осуществления одной из гипотез $B_1, B_2 ... B_n$. Как определить вероятность того, что имела место та или иная гипотеза?

При условии, что событие А уже произошло, вероятности гипотез *переоцениваются* по формулам, которые получили фамилию английского священника Томаса Байеса:

$P_A(B_1) = \dfrac{P(B_1) \cdot P_{B_1}(A)}{P(A)}$ – вероятность того, что имела место гипотеза B_1;

$P_A(B_2) = \dfrac{P(B_2) \cdot P_{B_2}(A)}{P(A)}$ – вероятность того, что имела место гипотеза B_2;

$P_A(B_3) = \dfrac{P(B_3) \cdot P_{B_3}(A)}{P(A)}$ – вероятность того, что имела место гипотеза B_3;

$P_A(B_n) = \dfrac{P(B_n) \cdot P_{B_n}(A)}{P(A)}$ – вероятность того, что имела место гипотеза B_n.

Задача

На склад поступило 2 партии изделий: первая – 4000 штук, вторая – 6000 штук. Средний процент нестандартных изделий в первой партии составляет 20%, а во второй – 10%. Наудачу взятое со склада изделие оказалось стандартным. Найти вероятность того, что оно: а) из первой партии, б) из второй партии.

Первая часть решения состоит в использовании формулы полной вероятности. Иными словами, вычисления проводятся в предположении, что испытание ещё не произведено и событие *«изделие оказалось стандартным»* пока не наступило.

Рассмотрим две гипотезы:

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

B_2 – наудачу взятое изделие будет из 2-й партии.

Всего: 4000 + 6000 = 10000 изделий на складе. Вероятности гипотез соответственно:

$$P(B_1) = \frac{4000}{10000} = 0{,}4; \quad P(B_2) = \frac{6000}{10000} = 0{,}6.$$

Контроль:

$$P(B_1) + P(B_2) = 0{,}4 + 0{,}6 = 1$$

Рассмотрим зависимое событие: A – наудачу взятое со склада изделие будет стандартным.

В первой партии 100% – 20% = 80% стандартных изделий, поэтому:

$$P_{B_1}(A) = \frac{80}{100} = 0{,}8$$

– вероятность того, что наудачу взятое на складе изделие будет стандартным при условии, что оно принадлежит 1-й партии.

Аналогично, во второй партии 100% – 10% = 90% стандартных изделий и

$$P_{B_2}(A) = \frac{90}{100} = 0{,}9$$

– вероятность того, что наудачу взятое на складе изделие будет стандартным при условии, что оно принадлежит 2-й партии.

По формуле полной вероятности:

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) = 0{,}4 \cdot 0{,}8 + 0{,}6 \cdot 0{,}9 = 0{,}32 + 0{,}54 = 0{,}86$$

– вероятность того, что наудачу взятое на складе изделие будет стандартным.

Часть вторая. Пусть наудачу взятое со склада изделие оказалось стандартным. Эта фраза прямо прописана в условии, и она констатирует тот факт, что событиеА произошло**.**

По формулам Байеса:

а) $P_A(B_1) = \dfrac{P(B_1) \cdot P_{B_1}(A)}{P(A)} = \dfrac{0{,}4 \cdot 0{,}8}{0{,}86} = \dfrac{0{,}32}{0{,}86} = \dfrac{32}{86} = \dfrac{16}{43} \approx 0{,}37$

– вероятность того, что выбранное стандартное изделие принадлежит 1-ой партии;

б) $P_A(B_2) = \dfrac{P(B_2) \cdot P_{B_2}(A)}{P(A)} = \dfrac{0{,}6 \cdot 0{,}9}{0{,}86} = \dfrac{0{,}54}{0{,}86} = \dfrac{54}{86} = \dfrac{27}{43} \approx 0{,}63$

– вероятность того, что выбранное стандартное изделие принадлежит 2-ой партии.

После *переоценки* гипотезы B_1, B_2 , разумеется, по-прежнему образуют полную группу:

$$P_A(B_1) + P_A(B_2) = \frac{16}{43} + \frac{27}{43} = 1 \ \textit{(проверка)}$$

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

§4. Формула Бернулли, Лапласа, Пуассона

Если осуществляется несколько испытаний, к тому же вероятность события А в каждом испытании не зависит от исходов других испытаний, то такие испытания носят название независимых относительно события А.

Событие А в различных независимых испытаниях может иметь или различные вероятности, или одну и ту же вероятность.

Под сложным событием понимают совмещение нескольких отдельных простых событий. Если проводится n независимых испытаний. В каждом из них событие А может появиться или не появиться. Если во всяком испытании вероятность события А одна и та же, равная р. Значит, вероятность того, что событие А не наступит в каждом испытании также постоянна, причем равна она q = 1—р.

$$P_k(n) = C_n^k \cdot p^k q^{n-k}$$ -формула Бернулли

Обычно формулу Бернулли применяют, если число испытаний не очень велико, а вероятность в одном испытании не очень мала.

Задача

Найти вероятность того, что при 10 бросках монеты орёл выпадет 3 раза.

Решение

Сначала немного порассуждаем: всего проводится 10 повторных независимых испытаний. Сколькими способами можно выбрать 3 испытания, в которых выпадет орёл?

$$C_{10}^3 = \frac{10!}{7! \cdot 3!} = \frac{8 \cdot 9 \cdot 10}{6} = 120$$ способами!

Используем формулу Бернулли:

$$P_n^m = C_n^m p^m q^{n-m}$$, в данном случае:

$n = 10$ – всего испытаний;

$m = 3$ – количество испытаний, в которых должен появиться орёл;

$p = \frac{1}{2}$ – вероятность появления орла в каждом испытании;

$q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$ – вероятность появления решки в каждом испытании.

Таким образом:

$P_{10}^{3} = C_{10}^{3} p^3 q^7 = 120 \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^7 = \frac{120}{1024} = 0{,}1171875$ – вероятность того, что при 10 бросках монеты орёл выпадет ровно 3 раза.

Ответ: $P_{10}^{3} \approx 0{,}12$

Локальная теорема Лапласа

Если вероятность p появления случайного события A в каждом испытании постоянна, то вероятность $P_n(m)$ того, что в n испытаниях событие A наступит ровно m раз, приближённо равна:

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{npq}} \cdot \varphi(x)$$, где $$x = \frac{m - np}{\sqrt{npq}}$$.

При этом, чем больше n, тем рассчитанная вероятность $P_n(m)$ будет лучше приближать точное значению P_n^m, полученное *(хотя бы гипотетически)* по формуле Бернулли. Рекомендуемое минимальное количество испытаний – примерно 50-100, в противном случае результат $P_n(m)$ может оказаться далёким от истины. Кроме того, локальная теорема Лапласа работает тем лучше, чем вероятность p ближе к 0,5, и наоборот – даёт существенную погрешность при значениях p, близких к нулю либо единице. По этой причине ещё одним критерием эффективного использования формулы

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x)$$

является выполнение неравенства $npq > 10$ *(*≈ 10*)*.

Так, например, если $n = 50,\ p = 0{,}5$, то $npq = 50 \cdot 0{,}5 \cdot 0{,}5 = 12{,}5 > 10$ и применение теоремы Лапласа для 50-ти испытаний оправдано. Но если n=50 и p=0,1 , то npq=50*0,1*0,9=4,5<10 и приближение $P_n(m)$ *(к точному значению* P_n^m *)* будет плохим.

О том, почему $P_n(m) \approx P_n^m$ и об особенной функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$ мы поговорим на уроке о нормальном распределении вероятностей, а пока нам по-требуется формально-вычислительная сторона вопроса. В частности, важным фактом является четность этой функции:

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

Оформим официальные отношения с нашим примером:

<u>Задача 1</u>

Монета подбрасывается 400 раз. Найти вероятность того, что орёл выпадет ровно:

а) 200 раз;

б) 225 раз.

Решение

С чего начать решение**?** Сначала распишем известные величины, чтобы они были перед глазами:

$n=400$ – общее количество независимых испытаний;

$p=0{,}5$ – вероятность выпадения орла в каждом броске;

$q=1-p=0{,}5$ – вероятность выпадения решки.

а) Найдём вероятность того, что в серии из 400 бросков орёл выпадет ровно $m=200$ раз. Ввиду большого количества испытаний используем локальную теорему Лапласа:

$$P_n(m)\approx\frac{1}{\sqrt{npq}}\cdot\varphi(x)\text{, где }\varphi(x)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}},\quad x=\frac{m-np}{\sqrt{npq}}.$$

На первом шаге вычислим требуемое значение аргумента:

$$x=\frac{m-np}{\sqrt{npq}}=\frac{200-400\cdot 0{,}5}{\sqrt{400\cdot 0{,}5\cdot 0{,}5}}=\frac{200-200}{\sqrt{100}}=\frac{0}{10}=0$$

Далее находим соответствующее значение функции: $\varphi(0)$. Это можно сделать несколькими способами. В первую очередь, конечно же, напрашиваются непосредственные вычисления:

$$\varphi(0)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{0^2}{2}}=\frac{1}{\sqrt{2\pi}}\cdot e^{0}=\frac{1}{\sqrt{2\pi}}\approx 0{,}3989$$

Округление проводят, как правило, до 4-х знаков после запятой.

Кроме того, существует *таблица значений функции* $\varphi(x)$, которая есть практически в любой книге по теории вероятностей, в частности, в учебном пособии *В.Е. Гмурмана*. На заключительном этапе применим формулу

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x)$$
:

$P_{400}(200) \approx \frac{1}{10} \cdot \varphi(0) \approx 0{,}1 \cdot 0{,}3989 = 0{,}03989$ – вероятность того, что при 400 бросках монеты орёл выпадет ровно 200 раз.

Как видите, полученный результат очень близок к точному значению

$P_{400}^{200} = 0{,}0398693019637926$, вычисленному по формуле Бернулли.

б) Найдём вероятность того, что в серии из 400 испытаний орёл выпадет ровно $m = 225$ раз. Используем локальную теорему Лапласа.

$$x = \frac{m - np}{\sqrt{npq}} = \frac{225 - 200}{10} = \frac{25}{10} = 2{,}5$$

$$\varphi(2{,}5) \approx 0{,}0175$$

$P_{400}(225) \approx \frac{1}{10} \cdot \varphi(2{,}5) \approx 0{,}1 \cdot 0{,}0175 = 0{,}00175$ – искомая вероятность.

Интегральная теорема Лапласа

Если вероятность Р появления случайного события А в каждом испытании постоянна, то вероятность $P_n(m_1 \le m \le m_2)$ того, что в n испытаниях событие А наступит **не менее** m_1 **и не более** m_2 **раз** *(от* m_1 *до* m_2 *раз включительно)*, приближённо равна:

$$P_n(m_1 \le m \le m_2) \approx \Phi(x_2) - \Phi(x_1),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-\frac{z^2}{2}} dz, \quad x_2 = \frac{m_2 - np}{\sqrt{npq}}, \quad x_1 = \frac{m_1 - np}{\sqrt{npq}}$

При этом количество испытаний, разумеется, тоже должно быть достаточно большими вероятность р не слишком мала/велика *(ориентировочно npq>10),* иначе приближение будет неважным либо плохим.

Формула Пуассона

Если вероятность события мала ($p \le 0{,}1$), а число испытаний велико ($n \ge 10$), то наиболее удобно применять формулу Пуассона:

$$P_n(k) = \lambda^k e^{-\lambda}/k!\ , \text{ где } \lambda = np$$

Глава 3.
Основные характеристики случайных величин

Величина называется случайной, если в результате испытания она примет значение, которое заранее неизвестно, которое зависит от случайных причин.

Случайные величины бывают дискретные, непрерывные.

§1. Числовые характеристики дискретной случайной величины

Числовыми характеристиками дискретной случайной величины являются:

математическое ожидание, дисперсия и среднеквадратическое отклонение.

Характеристикой среднего значения случайной величины служит математическое ожидание.

Математическим ожиданием дискретной случайной величины называют сумму произведений всех ее возможных значений на их вероятности:

$$\sum_{i=1}^{n} x_i \cdot p_i = x_1 \cdot p_1 + x_2 \cdot p_2 + \cdots + x_n \cdot p_n$$

Свойства математического ожидания:

1. Математическое ожидание постоянной величины равно самой постоянной:

$$M(C) = C$$

2. Постоянный можно выносить за знак математического ожидания:

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

3. Математическое ожидание произведения взаимно независимых случайных величин равно произведению математических ожиданий сомножителей:

$$M(X_1 \cdot X_2 \cdot ... \cdot X_n) = M(X_1) \cdot M(X_2) \cdot ... \cdot M(X_n)$$

4. Математическое ожидание суммы случайных величин равно сумме математических ожиданий слагаемых:

$$M(X_1 + X_2 + ... + X_n) = M(X_1) + M(X_2) + ... + M(X_n)$$

(для разности аналогично)

Характеристиками рассеяния возможных значений случайной величины вокруг математического ожидания служат, в частности, дисперсия и среднее квадратическое отклонение.

Дисперсией случайной величины X называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$$D(X) = M[X - M(X)]^2$$

Дисперсию удобно вычислять по формуле:

$$D(X) = M(X^2) - [M(X)]^2$$

Свойства дисперсии:

1. Дисперсия постоянной равна нулю:

$$D(C) = 0$$

2. Постоянный множитель можно выносить за знак дисперсии, предварительно возведя его в квадрат:

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

3. Дисперсия суммы (разности) независимых случайных величин равна сумме дисперсий слагаемых:

$$D(X_1 \pm X_2 \pm ... \pm X_n) = D(X_1) + D(X_2) + ... + D(X_n)$$

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

Средним квадратическим отклонением случайной величины называют квадратный корень из дисперсии:

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

Рассмотрим следующие задачи

1. Математическое ожидание и дисперсия случайной величины X соответственно равны 0,5 и 5. Найти математическое ожидание и дисперсию случайной величины $2X-3$.

Решение.

Согласно свойствам математического ожидания и дисперсии, получаем:

$$M(2X-3)=M(2X)+M(-3)=2M(X)-3=2\cdot\frac{1}{2}-3=1-3=-2$$

$$D(2X-3)=4\cdot D(X)=4\cdot5=20$$

2. Случайные величины X и Y независимы, причем $D(X)=3$ и $D(Y)=5$. Найти $D(Z)$, если $Z=4\cdot X-5\cdot Y+3$.

Решение.

На основании свойств дисперсии получаем:

$$D(Z)=D(4\cdot X-5\cdot Y+3)=16\cdot D(X)+25\cdot D(Y)=16\cdot3+25\cdot5=48+125=173$$

3. Закон распределения ДСВ X задан таблицей распределения

x_i	1	2	3	4
p_i	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{3}$	c

Найти: c, $M(X)$, $D(X)$, $\sigma(X)$, $P\{X<3\}$.

1) Так как $\sum_{i=1}^{4} p_i=1$, т.е. $\frac{1}{8}+\frac{1}{4}+\frac{1}{3}+c=1$, следовательно

$$c=1-\frac{1}{8}-\frac{1}{4}-\frac{1}{3}=\frac{24-3-6-8}{24}=\frac{7}{24}$$

Т.о. закон распределения примет вид

x_i	1	2	3	4
p_i	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{7}{24}$

$$M(X)=\sum_{i=1}^{4} x_i\cdot p_i=1\cdot\frac{1}{8}+2\cdot\frac{1}{4}+3\cdot\frac{1}{3}+4\cdot\frac{7}{24}=\frac{1}{8}+\frac{1}{2}+1+\frac{7}{6}=\frac{3+12+24+28}{24}=\frac{67}{24};$$

2) Для вычисления дисперсии воспользуемся формулой:

$$D(X) = M(X^2) - [M(X)]^2$$

Сначала найдем математическое ожидание ДСВ X^2 для этого составим закон распределения этой СВ. Напоминаю, что для этого необходимо каждое значение ДСВ X возвести в квадрат, а вероятности оставляем прежними. При одинаковых значениях ДСВ вероятности складываем.

$$M(X^2) = 1\cdot\frac{1}{8} + 4\cdot\frac{1}{4} + 9\cdot\frac{1}{3} + 16\cdot\frac{7}{24} = \frac{1}{8} + 1 + 3 + \frac{14}{3} = \frac{3+96+112}{24} = \frac{211}{24};$$

$$D(X) = \frac{211}{24} - \left(\frac{67}{24}\right)^2 = \frac{24\cdot 211 - 67^2}{24^2} = \frac{5064-4489}{576} = \frac{575}{576};$$

3) Найдем среднее квадратическое отклонение:

$$\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{575}{576}} = \frac{5\sqrt{23}}{24}$$

4) $P\{X<3\} = P\{X=1\} + P\{X=2\} = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}$

4. Функция распределения ДСВ X имеет вид

$$F(x) = \begin{cases} 0, & x \le 0 \\ 0{,}2, & 0 < x \le 1, \\ 0{,}6, & 1 < x \le 2 \\ 0{,}9, & 2 < x \le 3 \\ 1, & x > 3 \end{cases}$$

Найти: $M(X),\ M(X^2)\ D(X),\ \sigma(X)$.

Решение.

Составляем закон распределения ДСВ X (т.е. выполняем операцию обратную той, которую мы делали в предыдущей статье)

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

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

Составляем закон распределения ДСВ X^2

x_i^2	0	1	4	9
p_i	0,2	0,4	0,3	0,1

$M(X^2)=0\cdot0{,}2+1\cdot0{,}4+4\cdot0{,}3+9\cdot0{,}1=0{,}4+1{,}2+0{,}9=2{,}5$

$D(X)=M(X^2)-[M(X)]^2=2{,}5-1{,}3^2=2{,}5-1{,}69=0{,}81$

$\sigma(X)=\sqrt{0{,}81}=0{,}9$

5. Независимые случайные величины X и Y заданы таблицами распределения вероятностей

x_i^2	10	20
p_i	0,2	0,8

y_i^2	30	40	50
p_i	0,5	0,3	0,2

Найти $D(X+Y)$ двумя способами:

1. Составив предварительно таблицу распределения СВ $Z=X+Y$;

2. Используя правило сложения дисперсий.

Решение.

Составим таблицу распределения ДСВ $Z=X+Y$.

Найдем $z_{ij}=x_i+y_j$

10+30=40	20+30=50
10+40=50	20+40=60
10+50=60	20+50=70

Т.о. значения ДСВ Z таковы: $z_1=40,\ z_2=50,\ z_3=60,\ z_4=70$

Найдем соответствующие им вероятности:

$p_1=P\{Z=40\}=P\{X=10,Y=30\}=0{,}2\cdot0{,}5=0{,}1$

$p_2=P\{Z=50\}=P\{X=10,Y=40\}+P\{X=20,Y=30\}=0{,}2\cdot0{,}3+0{,}8\cdot0{,}5=0{,}06+0{,}4=0{,}46$

$p_3=P\{Z=60\}=P\{X=10,Y=50\}+P\{X=20,Y=40\}=0{,}2\cdot0{,}2+0{,}8\cdot0{,}3=0{,}04+0{,}24=0{,}28$

$p_4 = P\{Z = 70\} = P\{X = 20, Y = 50\} = 0{,}8 \cdot 0{,}2 = 0{,}16$

Получаем ряд распределения случайной величины Z

z_i^2	40	50	60	70
p_i	0,1	0,46	0,28	0,16

$$M(Z) = \sum_{i=1}^{4} z_i \cdot p_i = 40 \cdot 0{,}1 + 50 \cdot 0{,}46 + 60 \cdot 0{,}28 + 70 \cdot 0{,}16 = 4 + 23 + 16{,}8 + 11{,}2 = 55;$$

$$M(Z^2) = \sum_{i=1}^{4} z_i^2 \cdot p_i = 1600 \cdot 0{,}1 + 2500 \cdot 0{,}46 + 3600 \cdot 0{,}28 + 4900 \cdot 0{,}16 = 160 + 1150 + 1008 + 784 = 3102;$$

$D(Z) = M(Z^2) - [M(Z)]^2 = 3102 - 3025 = 77$

2. Используя правило сложения дисперсий: $D(Z) = D(X+Y) = D(X) + D(Y)$

$D(X) = M(X^2) - [M(X)]^2$

$M(X) = 10 \cdot 0{,}2 + 20 \cdot 0{,}8 = 2 + 16 = 18;$

$M(X^2) = 100 \cdot 0{,}2 + 400 \cdot 0{,}8 = 20 + 320 = 340;$

$M(Y) = 30 \cdot 0{,}5 + 40 \cdot 0{,}3 + 50 \cdot 0{,}2 = 15 + 12 + 10 = 37$

$M(Y^2) = 900 \cdot 0{,}5 + 1600 \cdot 0{,}3 + 2500 \cdot 0{,}2 = 450 + 480 + 500 = 1430$

$D(Y) = 1430 - 1369 = 61$

$D(Z) = 16 + 61 = 77$

§2. Числовые характеристики непрерывных случайных величин

Для непрерывных случайных величин, так же, как и для дискретных, используют понятия математического ожидания и дисперсии.

Математическим ожиданием *непрерывной случайной величины* X *называется значение интеграла:*

$$M(X) = M_X = \int_{-\infty}^{+\infty} x f(x) dx,$$

где $f(x)$ – плотность вероятности.

Дисперсией **непрерывной случайной величины** X **называется значение интеграла:**

$$D(X) = D_X = \int_{-\infty}^{+\infty} (x - M_X)^2 f(x) dx.$$

Среднее квадратичное отклонение *непрерывной случайной величины* X *вычисляется как корень квадратный из дисперсии:*

$$\sigma_X = \sqrt{D_X}.$$

Мода (Mo) непрерывной случайной величины X – это такое ее значение, которому соответствует ксимальное значение ее плотности вероятности.

Медианой (Me) непрерывной случайной величины X называется такое ее значение, которое определяется равенством:

$$P(X < Me(X)) = P(X > Me(X)).$$

Основные свойства математического ожидания и дисперсии непрерывных случайных величин остаются такими же, как и для дискретных случайных величин.

Начальные и центральные моменты для непрерывных случайных величин находятся по формулам:

$$\nu_k = M(x^k) = \int_{-\infty}^{+\infty} x^k f(x) dx$$

$$\mu_k = M(X - M(X))^k = \int_{-\infty}^{+\infty} (x - M_X)^k f(x) dx$$

§3. Основные примеры распределений непрерывной случайной величины

Равномерное распределение

Непрерывная случайная величина считается равномерно распределенной**,** если ее плотность вероятности имеет вид:

$$f(x) = \begin{cases} 0, (x \le x_1) \\ \dfrac{1}{x_2 - x_1}, (x_1 < x \le x_2) \\ 0, (x > x_2) \end{cases}$$

Математическое ожидание случайной величины, имеющей равномерное распре-деление:

$$M_X = \int_{-\infty}^{+\infty} xf(x)dx = \int_{x_1}^{x_2} \frac{x}{x_2 - x_1} dx = \frac{1}{x_2 - x_1} \frac{x^2}{2}\Bigg|_{x_1}^{x_2} = \frac{{x_2}^2 - {x_1}^2}{2(x_2 - x_1)} = \frac{x_2 + x_1}{2}$$

Дисперсия может быть вычислена следующим образом:

$$D_X = \int_{x_1}^{x_2} \frac{x^2}{x_2 - x_1} dx - {M_X}^2 = \frac{1}{x_2 - x_1} \frac{x^3}{3}\Bigg|_{x_1}^{x_2} - {M_X}^2 = \frac{{x_2}^3 - {x_1}^3}{3(x_2 - x_1)} - {M_X}^2 =$$

$$= \frac{{x_1}^2 + x_1 x_2 + {x_2}^2}{3} - \frac{(x_2 + x_1)^2}{4} = \frac{{x_1}^2 - 2x_1 x_2 + {x_2}^2}{12} = \frac{(x_1 - x_2)^2}{12}$$

Среднее квадратичное отклонение будет иметь вид:

$$\sigma_X = \frac{x_2 - x_1}{2\sqrt{3}}.$$

Показательное распределение

Показательным (экспоненциальным) распределением непрерывной случайной величины X называется такое распределение, которое описывается следующим выражением для плотности вероятности:

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

где λ – постоянная положительная величина.

Функция распределения вероятности в этом случае имеет вид:

$$F(x) = \begin{cases} 0, (x < 0) \\ 1 - e^{\lambda x}, (x \ge 0) \end{cases}$$

Математическое ожидание случайной величины, имеющей экспоненциальное распределение, получаем на основании общей формулы с учетом того, что $f(x) = 0$ при $x < 0$:

$$M_X = \int_0^{+\infty} x\lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} xe^{-\lambda x} dx.$$

Интегрируя это выражение по частям, находим:

$$M_X = \frac{1}{\lambda}.$$

Дисперсию для экспоненциального распределения можно получить, используя выражение:

$$D_X = \int_{-\infty}^{+\infty} x^2 f(x)dx - M_X{}^2.$$

Подставляя выражение для плотности вероятности, находим:

$$D_X = \lambda \int_{-\infty}^{+\infty} x^2 e^{-\lambda x} dx - M_X{}^2$$

Вычисляя интеграл по частям, получаем:

$$D_X = \frac{1}{\lambda^2}$$

Глава 4

От теории вероятностей к математической статистике

Теория вероятностей изучает математические законы распределения случайных событий, и фактически является теоретической базой для математической статистики. Но, если в теории вероятностей обычно распределение задано тем или иным образом, и требуется найти вероятности, числовые характеристики (например, математическое ожидание, дисперсию и т.п.), построить графики функции и плотности распределения, то в задачах математической статистики, напротив, известны данные (выборка), собранные по результатам какого-то эксперимента или наблюдения, по которым следует определить закон распределения, наиболее подходящий в данном случае, достоверную с некоторой долей вероятности информацию о том, какими могут быть математическое ожидание или среднеквадратическое отклонение величины и т.п.

Математическая статистика

Если говорить строго, то **математическая статистика** - это раздел математики, который изучает методы сбора, систематизации, обработки и использования статистических данных для получения научно обоснованных выводов и принятия на их основе решений.

Почему же для обработки простых наборов данных требуется целая наука? Потому, что эти данные, как бы мы не старались, никогда не являются точными, содержат случайные ошибки. Это могут быть и погрешности измерительных приборов, и ошибки человеческие (связанные с тем, кто проводит исследование и измерение), и неоднородность данных или, конечно, их недостаточность (невозможно изучить, например, всех коров в мире, чтобы делать выводы об их удоях;), или опросить всех избирателей чтобы сделать прогноз выигрыша для кандидата на выборах).

Обычно исследователь многократно повторяет (если это физически возможно), свой опыт, получает большое количество однотипных данных, которые теперь надо обработать и сделать весомые выводы, которые позволят не только продвинуться глубже в изучении предмета (будь то удои коров или политические предпочтения), но и сделать выводы, прогнозы, принять важные экономические решения и т.д.

Именно математическая статистика дает методы для обработки данных, алгоритмы для проверки статистических гипотез, критерии адекватности и значимости выбранной модели или закона, обоснованные границы точности для параметров распределения, которые мы можем получить исходя из наших данных и т.п.

§1. Генеральная совокупность

Основу статистического исследования составляет множество данных, полученных в результате измерения одного или нескольких признаков. Реально наблюдаемая совокупность объектов, статистически представленная рядом наблюдений $x_1, x_2, ..., x_n$ случайной величины X, является выборкой, а гипотетически существующая (домысливаемая) —генеральной совокупностью. Генеральная совокупность может быть конечной (число наблюдений **N = const**) или бесконечной (**N** = ∞), а выборка из генеральной совокупности — это всегда результат ограниченного ряда n наблюдений. Число наблюдений n, образующих выборку, называется объемом выборки. Если объем выборки n достаточно велик (**n** → ∞) выборка считается большой**,** в противном случае она называется выборкой ограниченного объема. Выборка считается малой, если при измерении одномерной случайной величины X объем выборки не превышает 30 (**n <= 30**), а при измерении одновременно нескольких (**k**) признаков в многомерном пространстве отношение **n** к **k** не превышает **10 (n/k < 10)**. Выборка образует вариационный ряд, если ее члены являются порядковыми статистиками, т. е. выборочные значения случайной величины **X** упорядочены по возрастанию (ранжированы), значения же признака называются вариантами**.**

Пример. Практически одна и та же случайно отобранная совокупность объектов — коммерческих банков одного административного округа Москвы, может рассматриваться как выборка из генеральной совокупности всех коммерческих банков этого округа, и как выборка из генеральной совокупности всех коммерческих банков Москвы, а также как выборка из коммерческих банков страны и т.д.

§2. Основные способы отбора выборки

Достоверность статистических выводов и содержательная интерпретация результатов зависит от репрезентативности выборки, т.е. полноты и адекватности представления свойств генеральной совокупности, по отношению к которой эту выборку можно считать представительной. Изучение статистических свойств совокупности можно организовать двумя способами: с помощью сплошного и не сплошного наблюдения. Сплошное наблюдение предусматривает обследование всех единиц изучаемой совокупности, а не сплошное (выборочное) наблюдение — только его части.

Пять основных способов организации выборочного наблюдения

1. Простой случайный отбор, при котором n объектов случайно извлекаются из генеральной совокупности Nобъектов (например с помощью таблицы или датчика случайных чисел), причем каждая из возможных выборок имеют равную вероятность. Такие выборки называются собственно-случайными.

2. Простой отбор с помощью регулярной процедуры осуществляется с помощью механической составляющей (например, даты, дня недели, номера квартиры, буквы алфавита и др.) и полученные таким способом выборки называются механическими.

3. Стратифицированный отбор заключается в том, что генеральная совокупность объема N подразделяется на подсовокупности или слои (страты) объема $N_1, N_2, ..., N_r$ так что $N_1 + N_2 + ... + N_r = N$. Страты представляют собой однородные объекты с точки зрения статистических характеристик (например, население делится на страты по возрастным группам или социальной принадлежности; предприятия - по отраслям). В этом случае выборки называются стратифицированными (иначе, расслоенными, типическими, районированными).

4. Методы серийного отбора используются для формирования серийных или гнездовых выборок. Они удобны в том случае, если необходимо обследовать сразу "блок" или серию объектов (например, партию товара, продукцию определенной серии или население при территориально-административном делении страны). Отбор серий можно осуществить собственно-случайным или механическим способом. При этом проводится сплошное обследование определенной партии товара, или целой территориальной единицы (жилого дома или квартала);

5. Комбинированный (ступенчатый) отбор может сочетать в себе сразу несколько способов отбора (например, стратифицированный и случайный или случайный и механический); такая выборка называется комбинированной.

Виды отбора

По виду различаются индивидуальный, групповой и комбинированный отбор. При индивидуальном отборе в выборочную совокупность отбираются отдельные единицы генеральной совокупности, при групповом отборе — качественно однородные группы (серии) единиц, **а** комбинированный отбор предполагает сочетание первого и второго видов.

По методу отбора различают повторную и бесповторную выборку.

Бесповторным называется отбор, при котором попавшая в выборку единица не возвращается в исходную совокупность и в дальнейшем выборе не участвует; при этом численность единиц генеральной совокупности **N** сокращается в процессе отбора. При повторном отборе попавшая в выборку единица после регистрации возвращается в генеральную совокупность и таким образом сохраняет равную возможность наряду с другими единицами быть использованной в дальнейшей процедуре отбора; при этом численность единиц генеральной совокупности **N** остается неизменной (метод в социально-экономических исследованиях применяется редко). Однако, при большом **N** **(N** → ∞**)** формулы для бесповторного отбора приближаются к аналогичным для повторного отбора и практически чаще используются последние (**N = const**).

§3. Основные характеристики параметров генеральной и выборочной совокупности

В основе статистических выводов проведенного исследования лежит распределение случайной величины X, наблюдаемые же значения **(**$x_1, x_2, \dots, x_n$**)** называются реализациями случайной величины **X** (n — объем выборки). Распределение случайной величины Xв генеральной совокупности носит теоретический, идеальный характер, а ее выборочный аналог является эмпирическим распределением. Некоторые теоретические распределения заданы аналитически, т.е. их параметры определяют значение функции распределения $F(x)$ в каждой точке пространства возможных значений случайной величины X. Для выборки же функцию распределения определить трудно, а иногда невозможно, поэтому параметры оценивают по эмпирическим данным, а затем их подставляют в аналитическое выражение, описывающее теоретическое распределение. При этом предположение (или гипотеза) о виде распределения может быть как статистически верным, так и ошибочным. Но в любом случае восстановленное по выборке эмпирическое распределение лишь грубо характеризует истинное. Важнейшими параметрами распределений являются математическое ожидание μ и дисперсия σ^2.

По своей природе распределения бывают непрерывными и дискретными. Наиболее известным непрерывным распределением является нормальное. Выборочными аналогами параметров μ иσ^2для него являются: среднее значение $\overline{x}$ и эмпирическая дисперсия s^2. Среди дискретных в социально-экономических исследованиях наиболее часто применяется альтернативное (дихотомическое**)** распределение. Параметр математического ожидания μ этого распределения выражает относительную величину (или долю) единиц совокупности, которые обладают изучаемым признаком x (она обозначена буквой p); доля совокупности, не обладающая этим признаком, обозначается буквой **q (q = 1 — p)**. Дисперсия же σ^2 альтернативного распределения также имеет эмпирический аналог s^2.

В зависимости от вида распределения и от способа отбора единиц совокупности по-разному вычисляются характеристики параметров распределения. Основные из них для теоретического и эмпирического распределений приведены в табл. 9.1.

Долей выборки $\mathbf{k_n}$ называется отношение числа единиц выборочной совокупности к числу единиц генеральной совокупности:

$$\mathbf{k_n = n/N}.$$

Выборочная доля w — это отношение единиц, обладающих изучаемым признаком **x** к объему выборки **n**:

$$\mathbf{w = n_n/n}.$$

Пример. В партии товара, содержащей 1000 ед., при 5% выборке доля выборки $\mathbf{k_n}$ в абсолютной величине составляет 50 ед. (n = N*0,05); если же в этой выборке обнаружено 2 бракованных изделия, то выборочная доля брака **w** составит 0,04 (w = 2/50 = 0,04 или 4%).

Так как выборочная совокупность отлична от генеральной, то возникают ошибки выборки.

Таблица 1. Основные параметры генеральной и выборочной совокупностей

Характеристика параметров распределения	Совокупность	
	генеральная	выборочная
Объем выборки	**N**	n
Альтернативный признак		
Численность единиц совокупности, обладающих признаком x	N_x	n_x
Доля единиц, обладающих изучаемым признаком x	$p = \frac{N_x}{N}$	$w = \frac{n_x}{n}$
Дисперсия	$\sigma^2 = p(1-p)$	$\sigma^2 = w(1-w)$
Среднее квадратическое отклонение	$\sigma = \sqrt{p(1-p)}$	$\sigma = \sqrt{w(1-w)}$
Количественный признак		
Среднее значение признака	$\mu = \frac{\sum x_i}{N}$	$\bar{x} = \frac{\sum x_i}{n}$
Дисперсия	$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N-1}$	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$
Среднее квадратическое отклонение	$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N-1}}$	$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

§4. Ошибки выборки

При любом статистическом наблюдении (сплошном и выборочном) могут встретиться ошибки двух видов: регистрации и репрезентативности. Ошибки регистрации могут
иметь случайный и систематический характер. Случайные ошибки складываются из множества различных неконтролируемых причин, носят непреднамеренный характер и обычно по совокупности уравновешивают друг друга (например, изменения показателей прибора при температурных колебаниях в помещении).

Систематические ошибки тенденциозны, так как нарушают правила отбора объектов в выборку (например, отклонения в измерениях при изменении настройки измерительного прибора).

Пример. Для оценки социального положения населения в городе предусмотрено обследовать 25% семей. Если при этом выбор каждой четвертой квартиры основан на ее номере, то существует опасность отобрать все квартиры только одного типа (например, однокомнатные), что обеспечит систематическую ошибку и исказит результаты; выбор же номера квартиры по жребию более предпочтителен, так как ошибка будет случайной.

Ошибки репрезентативности присущи только выборочному наблюдению, их невозможно избежать и они возникают в результате того, что выборочная совокупность не полностью воспроизводит генеральную. Значения показателей, получаемых по выборке, отличаются от показателей этих же величин в генеральной совокупности (или получаемых при сплошном наблюдении).

Ошибка выборочного наблюдения ε есть разность между значением параметра в генеральной совокупности и ее выборочным значением. Для среднего значения количественного признака она равна: $\varepsilon_{\overline{x}} = |\mu - \overline{x}|$, а для доли (альтернативного признака) — $\varepsilon_w = |p - w|$.

Ошибки выборки свойственны только выборочным наблюдениям. Чем больше эти ошибки, тем больше эмпирическое распределение отличается от теоретического. Параметры эмпирического распределения $\overline{x}$ и s^2 являются случайными величинами, следовательно, ошибки выборки также являются случайными величинами, могут принимать для разных выборок разные значения и поэтому принято вычислять **среднюю ошибку**.

Средняя ошибка выборки есть величина $m = \sqrt{\frac{\sigma^2}{n}}$, выражающая среднее квадратическое отклонение выборочной средней от математического ожидания. Эта величина при соблюдении принципа случайного отбора зависит прежде всего от объема выборки n и от степени варьирования признака: чем больше n и чем меньше вариация признака (следовательно, и значение σ^2), тем меньше величина средней ошибки выборки m. Соотношение между дисперсиями генеральной и выборочной совокупностей выражается формулой:

$$\sigma^2 = s^2 * \frac{n}{n-1}$$

т.е. при достаточно больших n можно считать, что $\sigma = s$. Средняя ошибка выборки показывает возможные отклонения параметра выборочной совокупности от параметра генеральной. В табл. 9.2 приведены выражения для вычисления средней ошибки m выборки при разных методах организации наблюдения.

Таблица 2. Средняя ошибка (m) выборочных средней и доли для разных видов выборки

Вид выборки	Отбор	
	Повторный	Бесповторный
Количественный признак		
Собственно-случайная	$m_{\bar{x}} = \sqrt{\frac{s^2}{n}}$	$m_{\bar{x}} = \sqrt{\frac{s^2(1-n/N)}{n}}$
Механическая	–	-"-
Типическая (стратифицированная)	$m_{\bar{x}} = \sqrt{\frac{\overline{s_i^2}}{n}}$	$m_{\bar{x}} = \sqrt{\frac{\overline{s^2_i}(1-n/N)}{n}}$
Серийная	$m_{\bar{x}} = \sqrt{\frac{\delta_x^2}{r}}$	$m_{\bar{x}} = \sqrt{\frac{\delta_x^2(1-r/R)}{r}}$
Альтернативный признак		
Собственно-случайная	$m_p = \sqrt{\frac{w(1-w)}{n}}$	$m_p = \sqrt{\frac{(w(1-w)(1-n/N)}{n}}$
Механическая	-	-"-
Типическая (стратифицированная)	$m_p = \sqrt{\frac{\overline{w_i(1-w_i)}}{n}}$	$m_p = \sqrt{\frac{\overline{w_i(1-w_i)}(1-n/N)}{n}}$
Серийная	$m_{\bar{x}} = \sqrt{\frac{\delta_w^2}{r}}$	$m_{\bar{x}} = \sqrt{\frac{\delta_w^2(1-r/R)}{r}}$

Где $\Delta \overline{s_i^2}$ - средняя из внутригрупповых выборочных дисперсий для непрерывного признака;

$\overline{w_i(1-w_i)}$ - средняя из внутригрупповых дисперсий доли;

r — число отобранных серий, R — общее число серий;

$$\delta_{\bar{x}}^2 = \frac{\sum(\bar{x}_i * - \bar{x})^2}{r},$$

где $\overline{x_i^*}$ — средняя i-й серии;

$\bar{x}$ — общая средняя по всей выборочной совокупности для непрерывного признака;

$$\delta_{\bar{x}}^2 = \frac{\sum(\bar{w}_i * - \bar{w})^2}{r},$$

где $\overline{w_i^*}$ — доля признака в i-й серии;

$\overline{w}$ — общая доля признака по всей выборочной совокупности.

Однако о величине средней ошибки m можно судить лишь с определенной, вероятностью P (P ≤ 1). Ляпунов А.М. доказал, что распределение выборочных средних $\overline{x}$, а следовательно, и их отклонений от генеральной средней, при достаточно большом числе n приближенно подчиняется нормальному закону распределения при условии, что генеральная совокупность обладает конечной средней и ограниченной дисперсией.

Математически это утверждение для средней выражается в виде:

$$P\{|\mu - \overline{x}| <= \Delta_x\} = \frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{-\frac{t^2}{2}} dt = \Phi(t), \ (1)$$

а для доли выражение (1) примет вид:

$$P\{|p-w| <= \Delta_w\} = \Phi(t), \ (2)$$

Где

$$\Delta_x = t \times m = t\sqrt{\frac{\sigma^2}{n}}$$

- есть **предельная ошибка выборки**, которая кратна величине средней ошибки выборки m**,** а коэффициент кратности t — есть критерий Стьюдента ("коэффициент доверия"), предложенный У.С. Госсетом (псевдоним "Student"); значения t для разного объема выборки n хранятся в специальной таблице.

Значения функции Ф(t) при некоторых значениях t равны:

$t=1$ $\Phi(t) = 0{,}683$	$t=1{,}5$ $\Phi(t) = 0{,}866$
$t=2$ $\Phi(t) = 0{,}954$	$t=2{,}5$ $\Phi(t) = 0{,}988$
$t=3$ $\Phi(t) = 0{,}997$	$t=3{,}5$ $\Phi(t) = 0{,}999$

Следовательно, выражение (3) может быть прочитано так: с вероятностью **P = 0,683 (68,3%)** можно утверждать, что разность между выборочной и генеральной средней не превысит одной величины средней ошибки **m (t = 1)**, с вероятностью **P = 0,954 (95,4%)** — что она не превысит величины двух средних ошибок **m (t = 2) ,** с вероятностью **P = 0,997 (99,7%)** — не превысит трех значений **m (t = 3) .** Таким образом, вероятность того, что эта разность превысит трехкратную величину средней ошибки определяет уровень ошибки и составляет не более **0,3%**.

В табл. 3 приведены формулы для вычисления предельной ошибки выборки.

Таблица 3. Предельная ошибка (D) выборки для средней доли (p) для разных видов выборочного наблюдения

N	Вид выборки	Отбор	
		повторный	бесповторный
1	2	3	4
Количественный признак			
1	Собственно-случайная	$t\sqrt{\frac{s^2}{n}}$	$t\sqrt{\frac{s^2(1-n/N)}{n}}$
2	Механическая ($n \to \infty$)	–	-"-
3	Типическая (стратифицированная)	$t\sqrt{\frac{\overline{s_i^2}}{n}}$	$t\sqrt{\frac{\overline{s^2{}_i}(1-n/N)}{n}}$
4	Серийная	$t\sqrt{\frac{\delta_x^2}{r}}$	$t\sqrt{\frac{\delta_x^2(1-r/R)}{r}}$
Альтернативный признак			
5	Собственно-случайная	$t\sqrt{\frac{w(1-w)}{n}}$	$t\sqrt{\frac{(w(1-w)(1-n/N)}{n}}$
6	Механическая ($n \to \infty$)	–	-"-
7	Типическая (стратифицированная)	$t\sqrt{\frac{\overline{w_i(1-w_i)}}{n}}$	$t\sqrt{\frac{\overline{w_i(1-w_i)}(1-n/N)}{n}}$
8	Серийная	$t\sqrt{\frac{\delta_w^2}{r}}$	$t\sqrt{\frac{\delta_w^2(1-r/R)}{r}}$

§5. Распространение выборочных результатов на генеральную совокупность

Конечной целью выборочного наблюдения является характеристика генеральной совокупности. При малых объемах выборки эмпирические оценки параметров ($\overline{x}$ и w) могут существенно отклоняться от их истинных значений (μ и p). Поэтому возникает необходимость установить границы, в пределах которых для выборочных значений параметров ($\overline{x}$ и w) лежат истинные значения (μ и p).

Доверительным интервалом какого-либо параметра θ генеральной совокупности называется случайная область значений этого параметра, которая с вероятностью близкой к 1 (**надежностью**) содержит истинное значение этого параметра.

Предельная ошибка выборки **Δ** позволяет определить предельные значения характеристик генеральной совокупности и их **доверительные интервалы**, которые равны:

для средней $\mu = \bar{x} \pm \Delta_{\bar{x}} = \bar{x} \pm t\, m_{\bar{x}}$, (4а)

для доли $p = w \pm \Delta_w = w \pm t\, m_w$. (4б)

Нижняя граница **доверительного интервала** получена путем вычитания **предельной ошибки** из выборочного среднего (доли), а верхняя — путем ее добавления.

Доверительный интервал для средней использует предельную ошибку выборки и для заданного уровня достоверности P определяется по формуле:

$\bar{x} - t\, m_{\bar{x}} < \mu$ – истинное значение средней $< \bar{x} + t\, m_{\bar{x}}$; (5а)

а для относительного параметра (доли):

$w - t\, m_p < p$ – истинное значение доли $< w + t\, m_p$. (5б)

Это означает, что с заданной вероятностью **P**, которая называется доверительным уровнем и однозначно определяется значением **t**, можно утверждать, что истинное значение средней лежит в пределах от

$$\bar{x} - t\, m_{\bar{x}} \text{ до } \bar{x} + t\, m_{\bar{x}},$$

а истинное значение доли w — в пределах от

$$w - t\, m_p \text{ до } w + t\, m_p.$$

При расчете доверительного интервала для трех стандартных доверительных уровней **P = 95%, P = 99% и P = 99,9%** значение t выбирается по таблице Стьюдента. Приложения в зависимости от числа степеней свободы $v = n - 1$. Если объем выборки достаточно велик, то соответствующие этим вероятностям значения **t** равны: **1,96, 2,58** и **3,29**. Таким образом, предельная ошибка выборки позволяет определить предельные значения характеристик генеральной совокупности и их доверительные интервалы:

Распространение результатов выборочного наблюдения на генеральную совокупность в социально-экономических исследованиях имеет свои особенности, так как требует полноты представительности всех ее типов и групп. Основой для возможности такого распространения является расчет относительной ошибки**:**

$$\Delta_{\%} = \frac{\Delta_{\bar{x}}}{\bar{x}} 100\% \text{ – для средней;}$$

$$\Delta_{\%} = \frac{\Delta_w}{\bar{p}} 100\% \text{ – для доли,}$$

где **$\Delta_{\%}$**- относительная предельная ошибка выборки; $\bar{x}$, $\bar{p}$.

Существуют два основных метода распространения выборочного наблюдения на генеральную совокупность**:** прямой пересчет и способ коэффициентов**.**

Сущность **прямого пересчета** заключается в умножении выборочного среднего значения !!\overline{x} на объем генеральной совокупности N.

Пример. Пусть среднее число детей ясельного возраста в городе оценено выборочным методом и составило $\overline{x} = 1.2$человека. Если в городе 1000 молодых семей, то число необходимых мест в муниципальных детских яслях получают умножением этой средней на численность генеральной совокупности N = 1000, т.е. составит 1200 мест.

Способ коэффициентов целесообразно использовать в случае, когда выборочное наблюдение проводится с целью уточнения данных сплошного наблюдения.

При этом используют формулу:

$$Y_1 = Y_0 \frac{y_1}{y_0},$$

где все переменные — это численность совокупности:

- Y_1 — с поправкой на недоучет,
- Y_0 - без этой поправки,
- y_0 — в контрольных точках
- y_1 — в тех же точках по данным контрольных мероприятий.

Необходимый объем выборки

При планировании выборочного наблюдения с заранее заданным значением допустимой ошибки выборки необходимо правильно оценить требуемый объем выборки**.** Этот объем может быть определен на основе допустимой ошибки при выборочном наблюдении исходя из заданной вероятности P, гарантирующей допустимую величину уровня ошибки (с учетом способа организации наблюдения). Формулы для определения необходимой численности выборки n легко получить непосредственно из формул предельной ошибки выборки. Так, из выражения для предельной ошибки:

$$\Delta = t\sqrt{\frac{\sigma^2}{n}}$$

непосредственно определяется объем выборки **n**:

$$n = \frac{t^2\sigma^2}{\Delta^2}$$

Эта формула показывает, что с уменьшением предельной ошибки выборки Δ существенно увеличивается требуемый объем выборки n, который пропорционален дисперсии σ^2 и квадрату критерия Стьюдента t.

Для конкретного способа организации наблюдения требуемый объем выборки n вычисляется согласно формулам, приведенным в табл. .4.

Таблица 4. Необходимый объем (n) выборки для разных видов организации выборочного наблюдения

	Вид выборки	Отбор	
		повторный	бесповторный
	1	2	3
	Количественный признак		
1	Собственно-случайная	$n = \frac{t^2 s^2}{\Delta_{\bar{x}}^2}$	$n = \frac{t^2 s^2 N}{\Delta_{\bar{x}}^2 N + t^2 s^2}$
2	Механическая ($n \to \infty$)	–	-"-
3	Типическая (стратифицированная)	$n = \frac{t^2 \bar{s}^2}{\Delta_{\bar{x}}^2}$	$n = \frac{t^2 \bar{s}^2 N}{\Delta_{\bar{x}}^2 N + t^2 \bar{s}^2}$
4	Серийная	$r = \frac{t^2 \delta^2}{\Delta_{\bar{x}}^2}$	$r = \frac{t^2 \delta^2 R}{\Delta_{\bar{x}}^2 R + t^2 \delta^2}$
	Альтернативный признак		
5	Собственно-случайная	$n = \frac{t^2 w(1-w)}{\Delta_w^2}$	$n = \frac{t^2 w(1-w) N}{\Delta_w^2 N + t^2 w(1-w)}$
6	Механическая ($n \to \infty$)	–	-"-
7	Типическая (стратифицированная)	$n = \frac{t^2 \overline{w(1-w)}}{\Delta^2{}_w}$	$n = \frac{t^2 \overline{w(1-w)N}}{\Delta_w^2 N + t^2 \overline{w(1-w)}}$
8	Серийная	$r = \frac{t^2 w_r(1-w_r)}{\Delta_w^2}$	$r = \frac{t^2 w_r(1-w_r) R}{\Delta_w^2 R + t^2 w_r(1-w_r)}$

§6. Практические примеры расчета

Пример 1. Вычисление среднего значения и доверительного интервала для непрерывного количественного признака.

Для оценки скорости расчета с кредиторами в банке проведена случайная выборка 10 платежных документов. Их значения оказались равными (в днях):

10; 3; 15; 15; 22; 7; 8; 1; 19; 20.

Необходимо с вероятностью **P = 0,954** определить предельную ошибку **Δ** выборочной средней и доверительные пределы среднего времени расчетов.

Решение. Среднее значение вычисляется по формуле из табл. 9.1 для выборочной совокупности

$$\bar{x} = \frac{\sum x_i}{n} = 12 \text{ дней.}$$

Дисперсия вычисляется по формуле из табл. 9.1.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = 53 \text{ дня.}$$

Средняя квадратическая погрешность $\sigma = 7.3$ дня.

Ошибка средней вычисляется по формуле:

$$m = \frac{\sigma}{\sqrt{n}} = 2{,}3,$$

т.е. среднее значение равно **x ± m = 12,0 ± 2,3 дней**.

Достоверность среднего составила

$$t = \frac{\bar{x}}{m} = \frac{12}{2{,}3} = 5{,}21.$$

Предельную ошибку вычислим по формуле из табл. 9.3 для повторного отбора, так как численность генеральной совокупности N неизвестна, и для **P = 0,954** уровня достоверности.

$$\Delta_{\bar{x}} = \pm t\, m_{\bar{x}} = 2 \times 2{,}3 = 4{,}6 \text{ дней,}$$

Таким образом, среднее значение равно \`x ± D = \`x ± 2m = 12,0 ± 4,6, т.е. его истинное значение лежит в пределах от 7,4 до16,6 дней.

Использование таблицы Стьюдента. Приложения позволяет заключить, что для n = 10 — 1 = 9 степеней свободы полученное значение достоверно с уровнем значимости a £ 0,001, т.е. полученное значение среднего достоверно отличается от 0.

Пример 2. Оценка вероятности (генеральной доли) р.

При механическом выборочном способе обследования социального положения 1000 семей выявлено, что доля малообеспеченных семей составила **w = 0,3 (30%)** (выборка была **2%**, т.е. **n/N = 0,02**). Необходимо с уровнем достоверности **р = 0,997** определить показатель **р** малообеспеченных семей во всем регионе.

Решение. По представленным значениям функции **Φ(t)** найдем для заданного уровня достоверности **P = 0,997** значение**t = 3** (см. формулу 3). Предельную ошибку доли **w** определим по формуле из табл. 9.3 для бесповторного отбора (механическая выборка всегда является бесповторной):

$$\Delta_w = t\sqrt{\frac{(w(1-w)(1-n/N)}{n}} = 3\sqrt{\frac{0{,}3(1-0{,}3)}{1000}(1-0{,}02)} = 0{,}014.$$

Предельная относительная ошибка выборки в **%** составит:

$$\Delta = \Delta_w/w = 0{,}014/0{,}3 = 4{,}7.$$

Вероятность (генеральная доля) малообеспеченных семей в регионе составит $\mathbf{p=w\pm\Delta_w}$, а доверительные пределы р вычисляются исходя из двойного неравенства:

$\mathbf{w - \Delta_w \leq p \leq w - \Delta_w}$, т.е. истинное значение р лежит в пределах:

0,3 — 0,014 < р <0,3 + 0,014, а именно от 28,6% до 31,4%.

Таким образом, с вероятностью 0,997 можно утверждать, что доля малообеспеченных семей среди всех семей региона составляет от 28,6% до 31,4%.

Пример 3. Вычисление среднего значения и доверительного интервала для дискретного признака, заданного интервальным рядом.

В табл. 5. задано распределение заявок на изготовление заказов по срокам их выполнения предприятием.

Таблица 5. Распределение наблюдений по срокам появления

Срок выполнения заявок (мес.)	Число наблюдений f_i(абсолютная частота)	Относительная частота p_i (%)	Середина интервала (градации) признака x_i
до 6	20	10	3
6-12	80	40	9
12-36	60	30	24
36-60	20	10	48
св.60	20	10	72
Всего	200	100%	

Решение. Средний срок выполнения заявок вычисляется по формуле:

$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i}$, где $\bar{x}$ – середина i-го интервала.

Средний срок составит:

$\bar{x}$ = (3*20 + 9*80 + 24*60 + 48*20 + 72*20)/200 = 23,1 мес.

Тот же ответ получим, если используем данные о p_i из предпоследней колонки табл. 9.5, используя формулу:

$\bar{x} = \sum x_i p_i$, $\bar{x} = 3\times10 + 9\times40 + 24\times30 + 48\times10 + 72\times10 = 23{,}1 \cong 23$ мес.

Заметим, что середина интервала для последней градации находится путем искусственного ее дополнения шириной интервала предыдущей градации равной 60 — 36 = 24 мес.

Дисперсия вычисляется по формуле

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1},$$

где **x_i**- середина интервального ряда.

Следовательно !!\sigma = \frac {20^2 + 14^2 + 1 + 25^2 + 49^2}{4}, а средняя квадратическая погрешность $\sigma = 30$.

Ошибка средней вычисляется по формуле $m = \sigma / \sqrt{n} = 13.4$ мес., т.е. среднее значение равно !!\overline{x} ± m = 23,1 ± 13,4.

Предельную ошибку вычислим по формуле из табл. 9.3 для повторного отбора, так как численность генеральной совокупности N неизвестна, для 0,954 уровня достоверности:

$$\Delta_{\bar{x}} = \pm\, t\, m_{\bar{x}} = 2\times13{,}4 = 26{,}8 \text{ дней},$$

Таким образом, среднее значение равно:

$$\bar{x} \pm \Delta_{\bar{x}} = 23{,}1 \pm 26{,}8,$$

т.е. его истинное значение лежит в пределах от 0 до 50 мес.

Пример 4. Для определения скорости расчетов с кредиторами N = 500 предприятий корпорации в коммерческом банке необходимо провести выборочное исследование методом случайного бесповторного отбора. Определить необходимый объем выборки n, чтобы с вероятностью P = 0,954 ошибка среднего значения выборки не превышала 3-х дней, если пробные оценки показали, что среднее квадратическое отклонение s составило 10 дней.

Решение. Для определения числа необходимых исследований n воспользуемся формулой для бесповторного отбора из табл. 9.4:

$$n=\frac{t^2 s^2 N}{\Delta_{\bar{x}}^2 N+t^2 s^2}$$

В ней значение t определяется из таблицы Стьюдента для уровня достоверности P = 0,954. Оно равно 2. Среднее квадратическое значение s = 10, объем генеральной совокупности N = 500, а предельная ошибка среднего значения $\mathbf{\Delta_x}$= 3. Подставляя эти значения в формулу, получим:

$$n=\frac{2^2\cdot 10^2\cdot 500}{3^2\cdot 500+2^2\cdot 10^2}=\frac{200000}{4900}=41$$

т.е. выборку достаточно составить из 41 предприятия, чтобы оценить требуемый параметр — скорость расчетов с кредиторами.

§7. Статистические оценки параметров распределения

Задача 1.

Из большой группы предприятий одной из отраслей промышленности случайным образом отобрано 30, по которым получены показатели основных фондов в млн. руб.: 2; 3; 2; 4; 5; 2; 3; 3; 6; 4; 5; 4; 6; 5; 3; 4; 2; 4; 3; 3; 5; 4; 6; 4; 5; 3; 4; 3; 2; 4.

1. Составить дискретное статистическое распределение выборки.
2. Найти объем выборки.
3. Составить распределение относительных частот.
4. Построить полигон частот.
5. Составить эмпирическую функцию распределения и построить ее график.
6. Найти несмещенные оценки числовых характеристик случайной величины.

Решение

1. Расположим различные значения признака в порядке их возрастания и под каждым из них запишем их частоты. Получим дискретное статистическое распределение выборки:

x_i	2	3	4	5	6
n_i	5	8	9	5	3

где x_i - варианты, n_i - частоты вариант x_i.

2. Сумма частот всех вариант должна быть равной объему выборки.

В данном примере объем выборки равен: n=5 + 8 + 9 + 5 + 3=30.

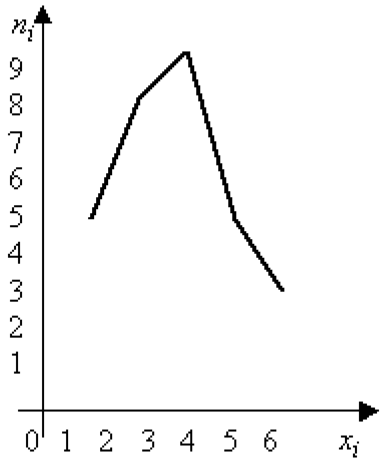
3. Найдем относительные частоты:

$$W_1 = \frac{5}{30} = \frac{1}{6};\ W_2 = \frac{8}{30} = \frac{4}{15};\ W_5 = \frac{3}{30} = \frac{1}{10}.\ W_3 = \frac{9}{30} = \frac{3}{10};\ W_4 = \frac{5}{30} = \frac{1}{6};$$

Запишем искомое распределение относительных частот

x_i	2	3	4	5	6
W_i	$\frac{1}{6}$	$\frac{4}{15}$	$\frac{3}{10}$	$\frac{1}{6}$	$\frac{1}{10}$

Контроль: $\frac{1}{6} + \frac{4}{15} + \frac{3}{10} + \frac{1}{6} + \frac{1}{10} = 1.$



Строим точки с координатами (x_i, n_i) и соединяем их последовательно отрезками. Полученная ломаная линия называется полигоном частот:

2. Согласно определению эмпирической функцией распределения называется функция вида

$$F^*(x) = \frac{n_x}{n},$$

3. где n – объем выборки; n_x - сумма частот вариант, меньших x.

Эмпирическая функция является оценкой функции распределения генеральной совокупности. Наименьшая варианта равна 2, поэтому при $x \le 2, n_x = 0$ и $F^*(x) = 0.$ Значение X<3, а именно, $X = x_1 = 2$ наблюдалось 5 раз. Тогда для $2 < x \le 3 \ n_x = 5$ и $F^*(x) = \frac{5}{30}.$ Значение X< 4, а именно, X=2, X=3, наблюдалось 5 + 8 =13 раз. Поэтому для $3 < x \le 4 \ n_x = 13$ и $F^*(x) = \frac{13}{30}.$

Аналогично рассуждая, получаем: для $4 < x \le 5 \ n_x = 5+8+9 = 22$ и $F^*(x) = \frac{22}{30},$

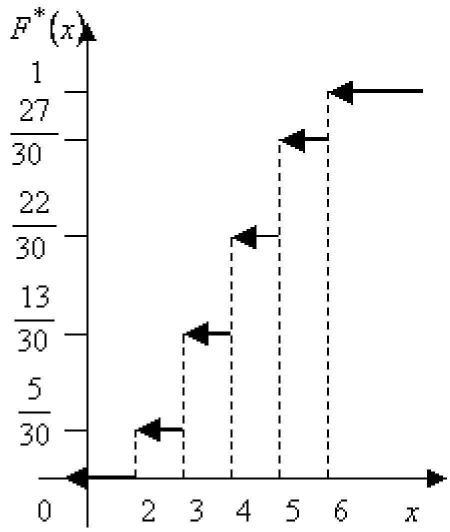
для $5 < x \le 6 \ n_x = 5+8+9+5 = 27$ и $F^*(x) = \frac{27}{30}$

и при $x > 6 \ n_x = 5+8+9+5+3 = 30$ и $F^*(x) = \frac{30}{30} = 1.$

Таким образом,

$$F^*(x) = \begin{cases} 0 \text{ при } & x \le 2; \\ \frac{5}{30} \text{ при } 2 < x \le 3; \\ \frac{13}{30} \text{ при } 3 < x \le 4; \\ \frac{22}{30} \text{ при } 4 < x \le 5; \\ \frac{27}{30} \text{ при } 5 < x \le 6; \\ 1 \text{ при } & x > 6. \end{cases}$$

График эмпирической функции имеет вид:



3. Несмещенной оценкой математического ожидания является средняя выборочная:

$$\bar{x}_B = \frac{\sum\limits_{i=1}^{k} x_i n_i}{n} = \frac{2\cdot 5+3\cdot 8+4\cdot 9+5\cdot 5+6\cdot 3}{30} = \frac{113}{30} \approx 3{,}77.$$

Несмещенная оценка дисперсии – исправления выборочная дисперсия:

$$s^2 = \frac{n}{n-1} D_B.$$

$$D_B = \overline{x_B^2} - \left(\bar{x}_B\right)^2 = \frac{2^2\cdot 5+3^2\cdot 8+4^2\cdot 9+5^2\cdot 5+6^2\cdot 3}{30} - \left(\frac{113}{30}\right)^2 \approx 1{,}42.$$

$$s^2 = \frac{30}{29}\cdot 1{,}42 \approx 1{,}47.$$

Задача 2.

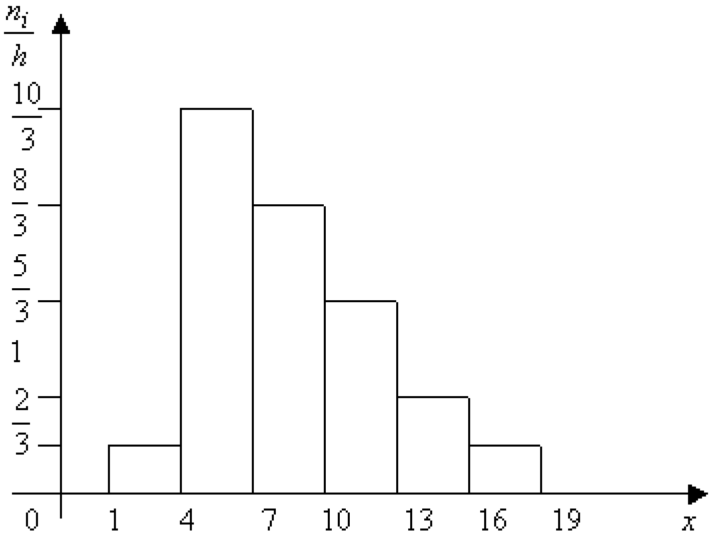
Выборочно обследование 30 предприятий машиностроительной промышленности по валовой продукции и получены следующие данные, в млн. руб.:

18,0; 12,0; 11,9; 1,9; 5,5; 14,6; 4,8; 5,6; 4,8; 10,9; 9,7; 7,2; 12,4; 7,6; 9,7; 11,2; 4,2; 4,9; 9,6; 3,2; 8,6; 4,6; 6,7; 8,4; 6,8; 6,9; 17,9; 9,6; 14,8; 15,8.

Составить интервальное распределение выборки с началом $x_0 = 1$ и длиной частичного интервала $h = 3$. Построить гистограмму частот. **Решение.** Для составления интервального распределения составим таблицу, в первой строке которой расположим в порядке возрастания интервалы, длина каждого из которых $h = 3$. Во второй строке запишем количество значений признака в выборке, попавших в этот интервал (т.е. сумму частот вариант, попавших в соответствующий интервал):

$(x_i; x_{i+1})$	1-4	4-7	7-10	10-13	13-16	16-19
n_i	2	10	8	5	3	2

Объем выборки $n = 2 + 10 + 8 + 5 + 3 + 2 = 30$.



Для построения гистограммы частот на оси абсцисс откладываем частичные интервалы, на каждом из них строим прямоугольники высотой $\frac{n_i}{h}$, где n_i - частота i-го частичного интервала, h – шаг (длина интервала), таким образом, гистограмма примет вид:

Указание. Для построения эмпирической функции распределения и нахождения точечных оценок ряда необходимо преобразовать его к дискретному виду по формуле

$$x_i^* = \frac{x_i + x_{i+1}}{2}.$$

Получим

x_i^*	2,5	5,5	8,5	11,5	14,5	17,5
n_i	2	10	8	5	3	2

Задача 3.

Из большой партии электроламп случайным образом отобрано 100. Средняя продолжительность горения ламп в выборке оказалась равной 1000 ч. Найти с надежностью $\gamma = 0{,}95$ доверительный интервал для средней продолжительности a горения ламп во всей партии, если известно, что среднее квадратическое отклонение продолжительности горения лампы $\sigma = 40$ ч

и продолжительность горения ламп распределена по нормальному закону.

Решение. По условию $\bar{x}_B = 1000,\ \gamma = 0{,}95,\ \sigma = 40.$

Для решения воспользуемся формулой

$$\bar{x}_B - \frac{\sigma t}{\sqrt{n}} < a < \bar{x}_B + \frac{\sigma t}{\sqrt{n}}$$

По приложению 3 находим *t* из условия:

$$\Phi(t) = \frac{\gamma}{2} = \frac{0{,}95}{2} = 0{,}475; \Rightarrow t = 1{,}96.$$

Тогда доверительный интервал:

$$1000 - \frac{1{,}96 \cdot 40}{\sqrt{100}} < a < 1000 + \frac{1{,}96 \cdot 40}{\sqrt{100}}$$
$$992{,}16 < a < 1007{,}84.$$

§8. Доверительный интервал для оценки истинного значения случайной величины

<u>Задача</u>

В результате 10 независимых измерений некоторой случайной величины X, выполненных с одинаковой точностью, получены опытные данные, приведенные в таблице.

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
6,9	7,3	7,1	9,5	9,7	7,9	7,6	9,1	6,6	9,9

Предполагая, что результаты измерений подчинены нормальному закону распределения вероятностей, оценить истинное значение X при помощи доверительного интервала, покрывающего истинное значение величины X с доверительной вероятностью 0,95.

Решение:

Поскольку в задаче имеется выборка малого объема, применим распределение Стьюдента.

Фактически требуется построить доверительный интервал для оценки математического ожидания *a* при неизвестном значении среднеквадратического отклонения из нормально распределенной генеральной совокупности.

Требуется отыскать такое число $t_{\alpha,n-1}$, для которого верно равенство

$$\boxed{P\left(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha,n-1}<a<\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha,n-1}\right)=1-\alpha}$$

В этой формуле:
$\overline{X}$ - выборочное среднее
S - стандартное (среднеквадратическое) отклонение
a - математическое ожидание
n - объем выборки (нашем случае 10)
α - величина, в сумме с доверительной вероятностью дающая 1 (в нашем случае 0,05).

Величину $t_{\alpha,n-1}$ (в нашем случае $t_{0{,}05{,}9}$) находим по таблицам распределения Стьюдента. Она равна 2,262.
Находим выборочное среднее как среднее арифметическое

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}x_i=8{,}16$$

Рассчитаем среднеквадратическое отклонение через исправленную выборочную дисперсию:

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2\approx 1{,}26^2$$

Тогда $\frac{S}{\sqrt{n}}t_{\alpha,n-1}\approx 0{,}903$

Получаем: $\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha,n-1}=7{,}257;\quad \overline{X}+\frac{S}{\sqrt{n}}t_{\alpha,n-1}=9{,}063$

Ответ: истинное значение случайной величины лежит в доверительном интервале (7,257; 9,063) с доверительной вероятностью 0,95.

Проверка гипотезы о распределении Пуассона

Задача

Отдел технического контроля проверил п = 500 партий однотипных изделий и установил, что число X нестандартных деталей в одной партии имеет эмпирическое распределение, приведенное в таблице.

xi	0	1	2	3	4	5
ni	194	186	88	26	5	1

x– число нестандартных изделий в одной партии, n – количество партий, содержащих x нестандартных изделий.

Требуется при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что случайная величина X (число нестандартных изделий в одной партии) распределена по закону Пуассона.

Решение:

Находим выборочную среднюю

$$\bar{x}_B = \frac{1}{n}\sum_{i=0}^{5} n_i x_i \approx 0{,}93$$

В качестве оценки параметра l распределения Пуассона

$$P_n(x_i) = \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}$$

выберем полученное значение выборочного среднего

$$\lambda = 0{,}93.$$

Расчет теоретических частот ведем по формуле

$$nt_i = 500 \cdot P(x_i)$$

Расчетная таблица значений.

x(i)	n(i)	P(xi)	nt(i)	n(i)	nt(i)	n(i)-nt(i)	(n(i)-nt(i))^2	(n(i)-nt(i))^2/nt(i)
0	194	0,395	197,277	194	197,277	-3,277	10,738	0,0544
1	186	0,367	183,467	186	183,467	2,533	6,414	0,0350
2	88	0,171	85,312	88	85,312	2,688	7,223	0,0847
3	26	0,053	26,447	26	26,447	-0,447	0,200	0,0075
4	5	0,012	6,149	6	7,293	-1,293	1,671	0,2291
5	1	0,002	1,144					
Сумма	**500**	**0,9996**	**499,796**					**0,4107**

Параметры рассчитаны автоматически.

Малочисленные частоты $n_i \le 5$ можно объединить. Также объединяются и соответствующие им теоретические частоты.

Получили:

$$\chi^2_{набл} = 0{,}4107$$

Число степеней свободы k = s – r – 1, т.к. проверяется гипотеза о распределении Пуассона (т.е. проверяется один параметр), то r = 1, k = s – 2 = 3 (s = 5, т.к. после исключения малочисленных частот в таблице осталось 5 строк) По таблице получаем:

$$\chi^2_{кр}(0{,}05;3) = 7{,}8$$

Ответ: поскольку $\chi^2_{набл.} \prec \chi^2_{кр}$, гипотеза о том, что случайная величина распределена по закону Пуассона может быть принята.

§9.Критерий Пирсона. Проверка гипотезы о нормальном распределении.

Задача 1.

Используя критерий Пирсона, при уровне значимости *a* = 0,05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности *X* с эмпирическим распределением выборки объема *n* = 200.

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Решение.

1. Вычислим

$$\bar{x}_B = \frac{\sum\limits_{i=1}^{k} x_i n_i}{n} = 12{,}63$$

и выборочное среднее квадратическое отклонение

$$\sigma_B = \sqrt{\overline{x_B^2} - \left(\bar{x}_B\right)^2} = 4{,}695.$$

2. Вычислим теоретические частоты учитывая, что *n* = 200, *h* = 2, σ_B = 4,695, по формуле

$$n_i = \frac{200 \cdot 2}{4{,}695} \cdot \varphi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right) = 85{,}2 \cdot \varphi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right).$$

Составим расчетную таблицу (значения функции *j*(*x*) приведены в приложении 1).

i	x_i	$u_i = \frac{(x_i - \bar{x}_B)}{\sigma_B}$	$\varphi(u_i)$	$n_i' = \frac{n \cdot h}{\sigma_B} \cdot \varphi(u_i)$
1	5	-1,62	0,1074	9,1
2	7	-1,20	0,1942	16,5
3	9	-0,77	0,2966	25,3
4	11	-0,35	0,3752	32,0
5	13	0,08	0,3977	33,9
6	15	0,51	0,3503	29,8
7	17	0,93	0,2589	22,0
8	19	1,36	0,1582	13,5
9	21	1,78	0,0818	7,0

3. Сравним эмпирические и теоретические частоты. Составим расчетную таблицу, из которой найдем наблюдаемое значение критерия $\chi^2_{набл} = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'}$:

i	n_i	n_i'	$\lvert n_i - n_i' \rvert$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
1	15	9,1	5,9	34,81	3,8
2	26	16,5	9,5	90,25	5,5
3	25	25,3	0,3	0,09	0,0
4	30	32,0	2,0	4,0	0,1
5	26	33,9	7,9	62,41	1,8
6	21	29,8	8,8	77,44	2,6
7	24	22,0	2,0	4,0	0,2
8	20	13,5	6,5	42,25	3,1
9	13	7,0	6,0	36,0	5,1
Сумма	200				$\chi^2_{набл}$ =22,2

По таблице критических точек распределения χ^2 (приложение 6), по уровню значимости *a* = 0,05 и числу степеней свободы *k* = *s* – 3 = 9 – 3 = 6 находим критическую точку правосторонней критической области $\chi^2_{кр}$ (0,05; 6) = 12,6.

Так как $\chi^2_{набл}$ =22,2 > $\chi^2_{кр}$ = 12,6, гипотезу о нормальном распределении генеральной совокупности отвергаем. Другими словами, эмпирические и теоретические частоты различаются значимо.

Задача2

Представлены статистические данные.
Результаты измерений диаметров n = 200 валков после шлифовки обобщены в табл. (мм):

Таблица Частотный вариационный ряд диаметров валков

i	1	2	3	4	5	6	7	8
xi, мм	6,68	6,69	6,7	6,71	6,72	6,73	6,74	6,75
ni	2	3	12	6	11	14	30	25

i	9	10	11	12	13	14	15	16
xi, мм	6,76	6,77	6,78	6,79	6,8	6,81	6,82	6,83
ni	27	31	14	8	5	6	5	1

Требуется:
1) составить дискретный вариационный ряд, при необходимости упорядочив его;
2) определить основные числовые характеристики ряда;
3) дать графическое представление ряда в виде полигона (гистограммы) распределения;
4) построить теоретическую кривую нормального распределения и проверить соответствие эмпирического и теоретического распределений по критерию Пирсона. При проверке статистической гипотезы о виде распределения принять уровень значимости а = 0,05

Решение: Основные числовые характеристики данного вариационного ряда найдем по определению. Средний диаметр валков равен (мм):

$$x\text{ср} = \frac{1}{n}\sum_{k=1}^{16} n_i \cdot x_k = 6{,}753;$$

исправленная дисперсия (мм2):

$$D = \frac{1}{n-1}\sum_{k=1}^{16} n_i \cdot (x_k - x_{cp})^2 = 0{,}0009166;$$

исправленное среднее квадратическое (стандартное) отклонение (мм):

$$s = \sqrt{D} = 0{,}03028.$$

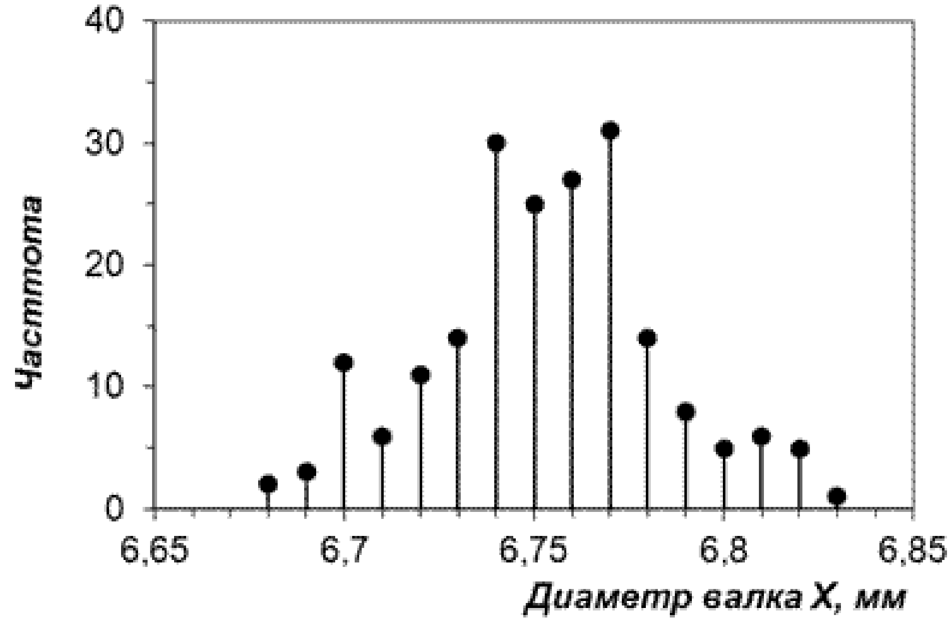


Рис. Частотное распределение диаметров валков

Исходное («сырое») частотное распределение вариационного ряда, т.е. соответствие $ni(xi)$, отличается довольное большим разбросом значений ni относительно некоторой гипотетической «усредняющей» кривой (рис.). В этом случае предпочтительно построить и анализировать интервальный вариационный ряд, объединяя частоты для диаметров, попадающих в соответствующие интервалы.

Число интервальных групп K определим по формуле Стерджесса:

K = 1 + log2n = 1 + 3,322lg$n$, где n = 200 – объем выборки. В нашем случае K = 1 + 3,322×lg200 = 1 + 3,322×2,301 = 8,644 » 8. Ширина интервала равна (6,83 – 6,68)/8 = 0,01875 » 0,02 мм. Интервальный вариационный ряд представлен в табл.

Таблица Частотный интервальный вариационный ряд диаметров валков

k	1	2	3	4	5	6	7	8
xk, мм	6,68 – 6,70	6,70 – 6,72	6,72 – 6,74	6,74 – 6,76	6,76 – 6,78	6,78 – 6,80	6,80 – 6,82	6,82 – 6,84
nk	5	18	25	55	58	22	11	6

Интервальный ряд может быть наглядно представлен в виде гистограммы частотного распределения.

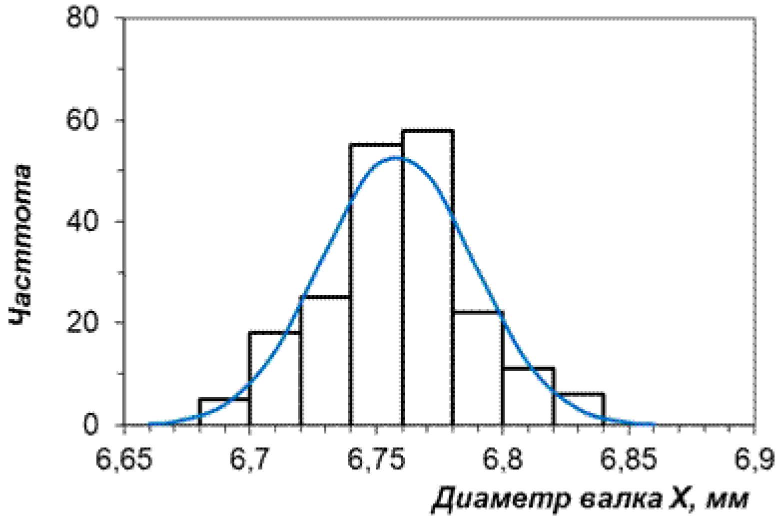


Рис. Частотное распределение диаметров валков. Сплошная линия – сглаживающая нормальная кривая

Вид гистограммы позволяет сделать предположение о том, что распределение диаметров валков подчиняется нормальному закону, согласно которому теоретические частоты могут быть найдены как

nk, теор = $n \times N(a;\ s;\ xk) \times Dxk$,

где, в свою очередь, сглаживающая гауссова кривая нормального распределения определяется выражением:

$$N(a;\ s;\ xk) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_k - a)^2}{2\sigma^2}\right\}.$$

В этих выражениях xk – центры интервалов в частотном интервальном вариационном ряде.

Например, $x1 = (6{,}68 + 6{,}70)/2 = 6{,}69$. В качестве оценок центра a и параметра s гауссовой кривой можно принять: $a = x$ср.

Из рис. видно, что гауссова кривая нормального распределения в целом соответствует эмпирическому интервальному распределению. Однако следует удостовериться в статистической значимости этого соответствия. Используем для проверки соответствия эмпирического распределения эмпирическому критерий согласия Пирсона c2 [2-4]. Для этого следует вычислить эмпирическое значение критерия как сумму

$$\chi^2 = \sum_{k=1}^{8} \frac{(n_k - n_{k,meop})^2}{n_{k,meop}},$$

где *nk* и *nk*,теор – эмпирические и теоретические (нормальные) частоты, соответственно. Результаты расчетов удобно представить в табличном виде:

Таблица Вычисления критерия Пирсона

[*xk*, *xk*+1), мм	*xk*, мм	*nk*	*nk*,теор	$\frac{(n_k - n_{k,meop})^2}{n_{k,meop}}$
6,68 – 6,70	6,69	5	4,00	0,25
6,70 – 6,72	6,71	18	14,57	0,81
6,72 – 6,74	6,73	25	34,09	2,42
6,74 – 6,76	6,75	55	51,15	0,29
6,76 – 6,78	6,77	58	49,26	1,55
6,78 – 6,80	6,79	22	30,44	2,34
6,80 – 6,82	6,81	11	12,07	0,09
6,82 – 6,84	6,83	6	3,07	2,80
			с2эмп	10,55

Критическое значение критерия найдем по таблице Пирсона [2, 3] для уровня значимости a = 0,05 и числа степеней свободы *d.f.* = *K* – 1 – *r*, где *K* = 8 – число интервалов интервального вариационного ряда; *r* = 2 – число параметров теоретического распределения, оцененных на основании данных выборки (в данном случае, – параметры *a* и s). Таким образом, *d.f.* = 5. Критическое значение критерия Пирсона есть χ^2крит(a; *d.f.*) = 11,1. Так как с2эмп < с2крит, заключаем, что согласие между эмпирическим и теоретическим нормальным распределением является статистическим значимым. Иными словами, теоретическое нормальное распределение удовлетворительно описывает эмпирические данные.

Задача3

Коробки с шоколадом упаковываются автоматически. По схеме собственно-случайной бесповторной выборки взято 130 из 2000 упаковок, содержащихся в партии, и получены следующие данные об их весе:

Вес упаковки (гр.)	Менее 975	975-1000	1000-1025	1025-1050	Более 1050	Всего
Число упаковок	6	38	44	34	8	130

Требуется используя критерий $\chi^2 -$ Пирсона при уровне значимости a=0,05 проверить гипотезу о том, что случайная величина X – вес упаковок – распределена по нормальному закону. Построить на одном графике гистограмму эмпирического распределения и соответствующую нормальную кривую.

Решение

$$\bar{x} = \mathbf{1012{,}5}$$

$$s^2 = \mathbf{615{,}3846}$$

<u>Примечание:</u>

В принципе в качестве дисперсии нормального закона распределения следует взять исправленную выборочную дисперсию. Но т.к. количество наблюдений – 130 достаточно велико, то подойдет и “обычная” s^2. Таким образом, теоретическое нормальное распределение имеет вид:

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

подставляем a = **1012,5** σ^2 = **615,3846** **σ = 24,8069**

$$f(x) = 0{,}0160819 \cdot e^{-0{,}0008125(x-1012{,}5)^2}$$

Для расчета вероятностей *pi*попадания случайной величины в интервал [*xi ; xi+1*] используем функцию Лапласа:

$$p_i(x_i \le X \le x_{i+1}) = \frac{1}{2}\left[\Phi\left(\frac{x_{i+1}-a}{\sigma}\right) - \Phi\left(\frac{x_i - a}{\sigma}\right)\right]$$

$$p_i(x_i \le X \le x_{i+1}) \approx \frac{1}{2}\left[\Phi\left(\frac{x_{i+1}-1012{,}5}{24{,}8069}\right) - \Phi\left(\frac{x_i - 1012{,}5}{24{,}8069}\right)\right]$$

в нашем случае получаем:

$$p_i(950 \le X \le 975) \approx \frac{1}{2}[\Phi(-1{,}51) - \Phi(-2{,}52)] = \frac{1}{2}[0{,}9883 - 0{,}869] = 0{,}0597$$

$$p_i(975 \le X \le 1000) \approx \frac{1}{2}[\Phi(-0{,}5) - \Phi(-1{,}51)] = \frac{1}{2}[0{,}869 - 0{,}3829] = 0{,}2431$$

$$p_i(1000 \le X \le 1025) \approx \frac{1}{2}[\Phi(0{,}5) - \Phi(-0{,}5)] = \frac{1}{2}[0{,}3829 + 0{,}3829] = 0{,}3829$$

$$p_i(1025 \le X \le 1050) \approx \frac{1}{2}[\Phi(1{,}51) - \Phi(0{,}5)] = \frac{1}{2}[0{,}869 - 0{,}3829] = 0{,}2431$$

$$p_i(1050 \le X \le 1075) \approx \frac{1}{2}[\Phi(2{,}52) - \Phi(1{,}51)] = \frac{1}{2}[0{,}9883 - 0{,}869] = 0{,}0597$$

Примечание: Такие симметричные вероятности получились из-за того, что по нашим начальным условиям выборочная средняя попала точно в середину среднего интервала выборки.

Составим таблицу:

	Интервал [*xi ; xi+1*]	Эмпирические частоты *ni*	*Вероятности pi*	*Теоретические частоты npi*	*(ni-npi)2*	$\frac{(n_i - np_i)^2}{np_i}$
	Менее 975	6	0,0597	7,761	3,101	0,3996
	975-1000	38	0,2431	31,603	40,922	1,2949
	1000-1025	44	0,3829	49,777	33,374	0,6705
	1025-1050	34	0,2431	31,603	5,746	0,1818
	Более 1050	8	0,0597	7,761	0,057	0,0073
		130	0,9885	128,5		$\chi^2 = 2{,}55$

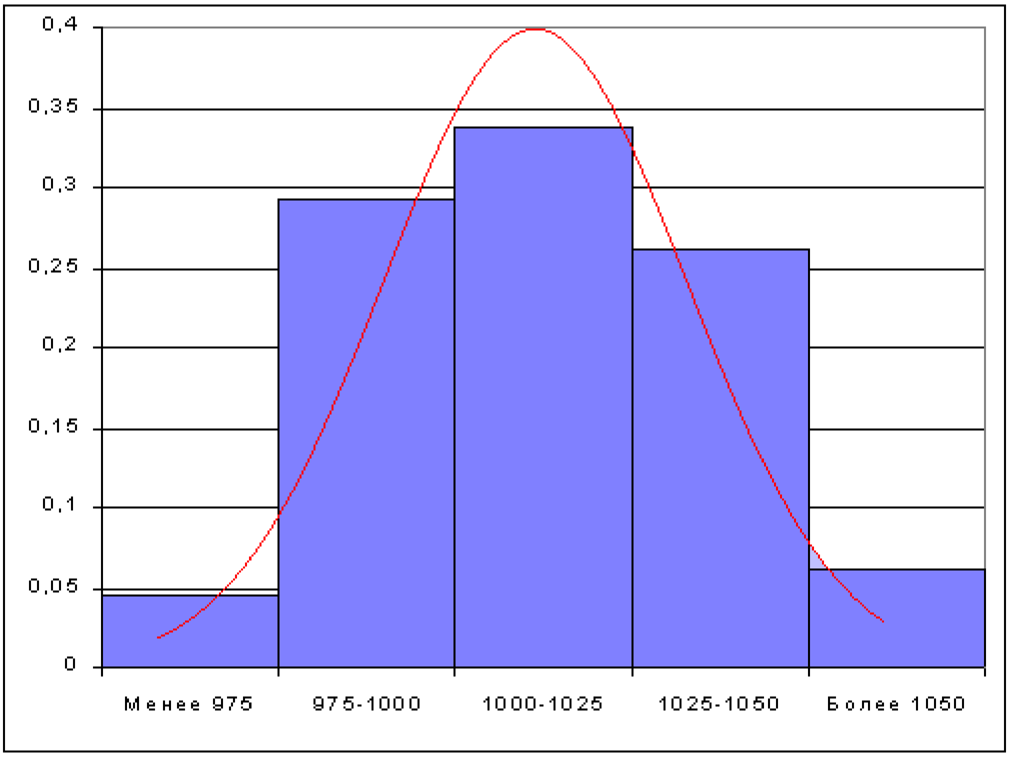
Итого, значение статистики $\chi^2 = 2{,}55$.

Определим количество степеней свободы по формуле: $k = m - r - 1$.
m – число интервалов (*m* = 5) *r*– число параметров закона распределения (в нормальном распределении *r* = 2) т.е. *k* = 2.

Соответствующее критическое значение статистики $\chi^2_{0,05;2} = 5{,}99$

Поскольку $\chi^2_{0,05;2} > \chi^2$, гипотеза о нормальном распределении с параметрами N(1012,5; 615,3846) согласуется с опытными данными.

Ниже показана гистограмма эмпирического распределения и соответствующая нормальная кривая.



Прямая регрессии

<u>Задача</u>

1) Вычислить групповые средние $\overline{x_i}$ и $\overline{y_i}$ и построить эмпирические линии регрессии.

2) Предполагая, что между переменными X и Y существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими линиями регрессии;

б) вычислить коэффициент корреляции на уровне значимости 0,05, оценить его значимость и сделать вывод о тесноте и направлении связи между переменными X и Y;

в) используя соответствующее уравнение регрессии, оценить средний объем выручки от продаж при ежемесячных затратах на рекламу в размере 2,4 тыс.руб.

Решение.

Находим групповые средние по формулам:

$$\overline{x_i} = \frac{\sum_{i=1}^{5} x_i n_{ij}}{n_i}; \quad \overline{y_i} = \frac{\sum_{i=1}^{6} y_i n_{ij}}{n_i};$$

x_i, y_i - середины соответствующих интервалов.

$x_j = [2.1, 2.3, 2.5, 2.7, 2.9]$

$y_i = [30, 34, 38, 42, 46, 50]$

Групповые средние:

$ycp_1 := 33.33333333$ $xcp_1 := 2.100000000$

$ycp_2 := 38.80000000$ $xcp_2 := 2.220000000$

$ycp_3 := 41.75000000$ $xcp_3 := 2.372727273$

$ycp_4 := 45.27272727$ $xcp_4 := 2.514285714$

$ycp_5 := 47.71428571$ $xcp_5 := 2.666666667$

$xcp_6 := 2.833333333$

Полученные по формулам значения заносим в таблицу:

X \ Y	28-32	32-36	36-40	40-44	44-48	48-52	Итого:	ycp[i]
2,0-2,2	2	3	1				**6**	33,33
2,2-2,4		1	6	3			**10**	38,8
2,4-2,6		1	3	8	4		**16**	41,75
2,6-2,8			1	2	6	2	**11**	45,27
2,8-3,0				1	2	4	**7**	47,71
Итого:	**2**	**5**	**11**	**14**	**12**	**6**	**50**	
xcp[i]	2,1	2,22	2,37	2,51	2,67	2,83		

Для нахождения уравнений регрессии вычисляем необходимые суммы:

$\sum_{i=1}^{5} x_i n_i =$ 2.1*6+2.3*10+2.5*16+2.7*11+2.9*7 = 125.6

$\sum_{i=1}^{5} x_i^2 n_i =$ $2.1^2\,6 + 2.3^2\,10 + 2.5^2\,16 + 2.7^2\,11 + 2.9^2\,7 = 318.42$

$\sum_{i=1}^{6} y_i n_i =$ 30*2+34*5+38*11+42*14+46*12+50*6 = 2088

$\sum_{i=1}^{6} y_i^2 n_i = 30^2\,2 + 34^2\,5 + 38^2\,11 + 42^2\,14 + 46^2\,12 + 50^2\,6 = 88552$

$$\sum_{i=1}^{5}\sum_{j=1}^{6} x_i y_j n_{ij} = 5295{,}6$$

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i n_i}{n} = \frac{125{,}6}{50} = 2{,}512$$

$$\bar{y} = \frac{\sum_{j=1}^{6} y_j n_j}{n} = \frac{2088}{50} = 41{,}76$$

$$s_x^2 = \frac{\sum_{i=1}^{5} x_i^2 n_i}{n} - \bar{x}^2 = \frac{318{,}42}{50} - 2{,}512^2 = 0{,}0583$$

$$s_y^2 = \frac{\sum_{j=1}^{6} y_j^2 n_j}{n} - \bar{y}^2 = \frac{88552}{50} - 41{,}76^2 = 27{,}142$$

$$\mu = \overline{xy} - \bar{x} \cdot \bar{y} = \frac{\sum_{i=1}^{5}\sum_{j=1}^{6} x_i y_j n_{ij}}{n} - \bar{x} \cdot \bar{y} = \frac{5295{,}6}{50} - 2{,}512 \cdot 41{,}76 = 1{,}011$$

$$b_{yx} = \frac{\mu}{s_x^2} = \frac{1{,}011}{0{,}0583} = 17{,}3524$$

$$b_{xy} = \frac{\mu}{s_y^2} = \frac{1{,}011}{27{,}142} = 0.0372$$

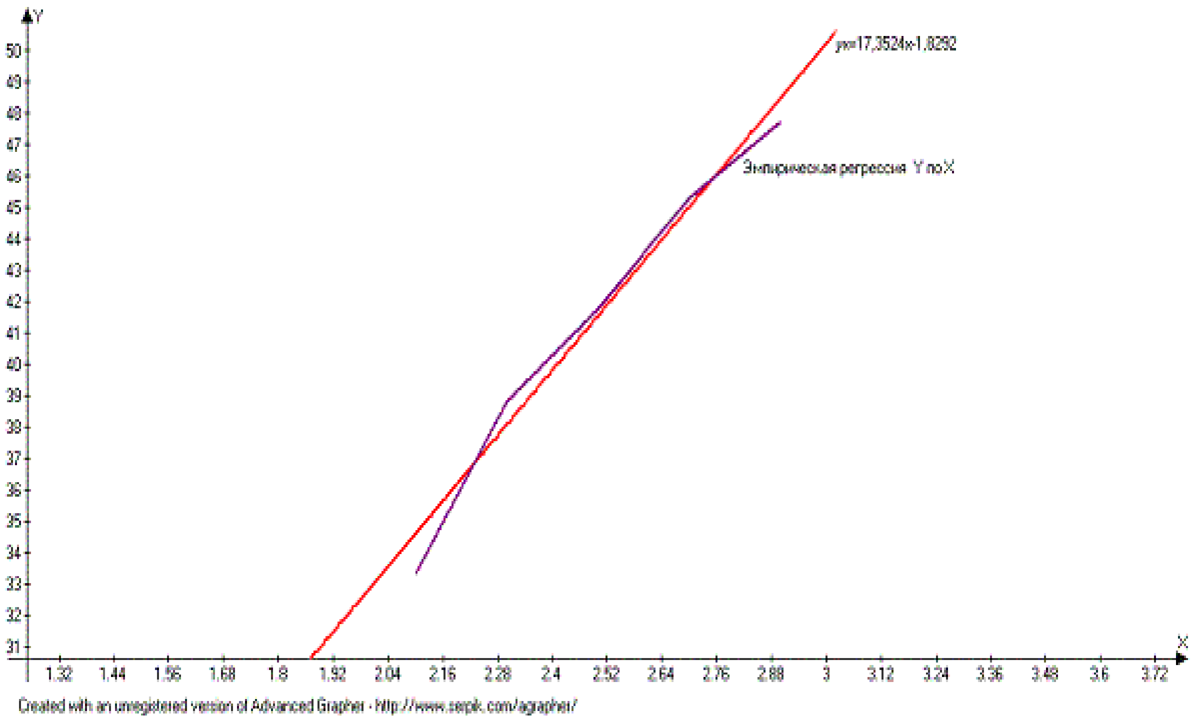
$$y_x - \bar{y} = b_{yx}(x - \bar{x})$$
$$y_x = b_{yx}x - b_{yx}\bar{x} + \bar{y}$$

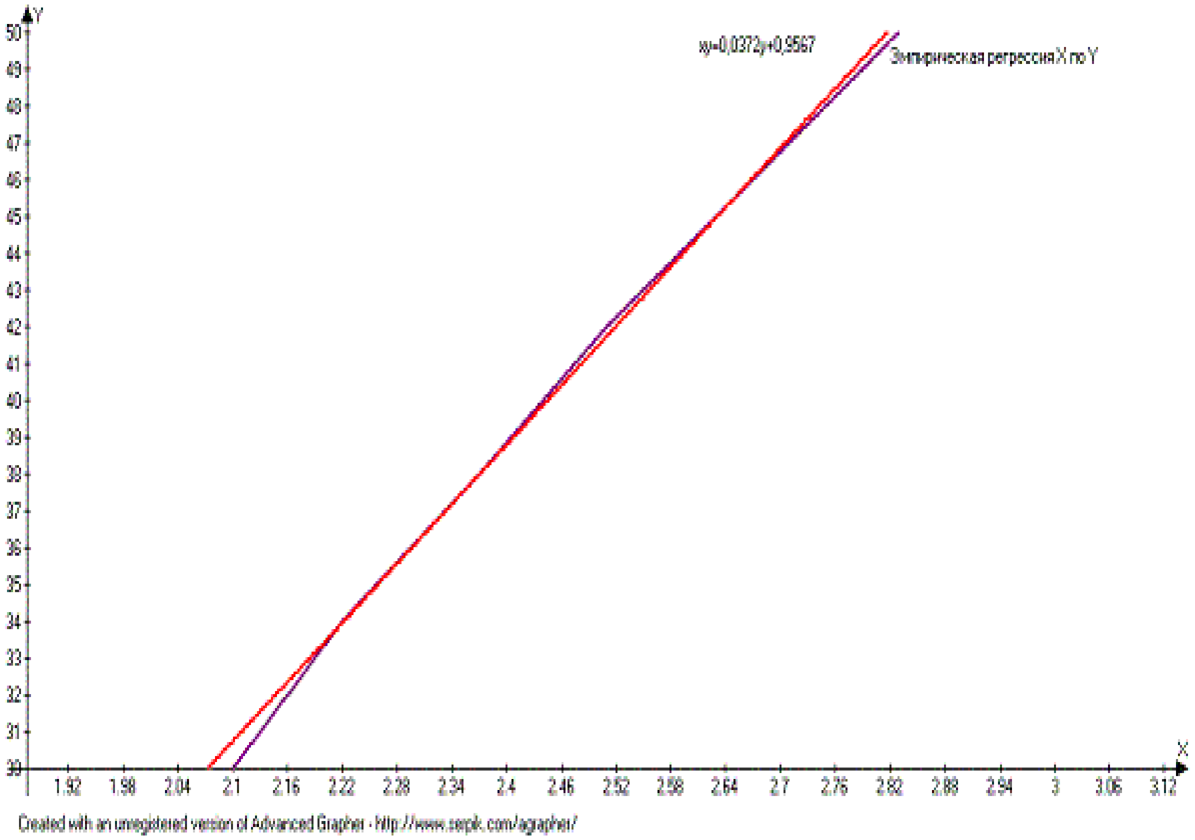
$$x_y - \bar{x} = b_{xy}(y - \bar{y})$$
$$x_y = b_{xy}y - b_{xy}\bar{y} + \bar{x}$$

Получаем искомые уравнения регрессии:

$$y_x = 17{,}3524x - 1{,}8292 \,, \quad x_y = 0{,}0372y + 0{,}9567$$

Ниже представлены графики полученных уравнений регрессии совместно с соответствующей эмпирической регрессией





Находим коэффициент корреляции $r = \pm\sqrt{b_{yx} b_{xy}}$ радикал берем со знаком + , т.к коэффициенты b_{yx} и b_{xy} положительны.

$$r = \sqrt{17{,}3524 \cdot 0{,}0372} = 0{,}8039$$

Оценим значимость коэффициента корреляции.

$$t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}=\frac{0{,}8039\sqrt{50-2}}{\sqrt{1-0{,}8039^2}}=9{,}364$$

По таблице критерия Стьюдента для уровня значимости 0,05 находим

$$t_{0{,}95;48}=2{,}01$$

Т.к. $t_{0{,}95;48}<t$, то коэффициент корреляции значимо отличается от нуля. Связь тесная и прямая.

По найденному уравнению регрессии находим:

$$y_{x=2{,}4}=17{,}3524\cdot 2{,}4-1{,}8292=39{,}82$$ млн.руб

Ответ:

Групповые средние:

$ycp_1 := 33.33333333$ $xcp_1 := 2.100000000$

$ycp_2 := 38.80000000$ $xcp_2 := 2.220000000$

$ycp_3 := 41.75000000$ $xcp_3 := 2.372727273$

$ycp_4 := 45.27272727$ $xcp_4 := 2.514285714$

$ycp_5 := 47.71428571$ $xcp_5 := 2.666666667$

$xcp_6 := 2.833333333$

Уравнения регрессии:

$$y_x=17{,}3524x-1{,}8292$$

$$x_y=0{,}0372y+0{,}9567$$

Коэффициент корреляции:

$$r=0{,}8039$$

$$y_{x=2{,}4}=39{,}82$$

§10. Задачи для самостоятельного решения

1. Перед Вами на 3 различных (груша, яблоко, банан) фрукта. Сколькими способами можно выбрать а) один фрукт, б) два фрукта, в) три фрукта, г) хотя бы один фрукт?

2. Предприятие может предоставить работу по одной специальности 4 женщинами, по другой - 6 мужчинам, по третьей - 3 работникам независимо от пола. Сколькими способами можно заполнить вакантные места, если имеются 14 претендентов: 6 женщин и 8 мужчин?

3. Группу из 20 студентов нужно разделить на 3 бригады, причем в первую бригаду должны входить 3 человека, во вторую — 5 и в третью — 12. Сколькими способами это можно сделать.

4. Цифры 1, 2, 3, …, 9, выписанные на отдельные карточки складывают в ящик и тщательно перемешивают. Наугад вынимают одну карточку. Найти вероятность того, что число, написанное на этой карточке: а) четное; б) двузначное.

5. На полке в случайном порядке расставлено 40 книг, среди которых находится трехтомник Пушкина. Найти вероятность того, что эти тома стоят в порядке возрастания номера слева направо, но не обязательно рядом.

6. С базы в магазин отправлено 4000 тщательно упакованных доброкачественных изделий. Вероятность того, что изделие повредится в пути, равна 0.0005. Найти вероятность того, что из 4000 изделий в магазин прибудут 3 испорченных изделия.

7. В магазине имеется 15 автомобилей определенной марки. Среди них 7 черного цвета, 6 серого и 2 белого. Представители фирмы обратились в магазин с предложением о продаже им 3 автомобилей этой марки, безразлично какого цвета. Составьте ряд распределения числа проданных автомобилей черного цвета при условии, что автомобили отбирались случайно.

8. В магазине продаются 5 отечественных и 3 импортных телевизора. Составить закон распределения случайной величины – числа импортных из четырех наудачу выбранных телевизоров. Найти функцию распределения этой случайной величины и построить ее график.

9. Случайная величина задана дифференциальной функцией распределения

$$f(x)=\begin{cases} 0 \ \text{при} \ x\le\pi, \\ -\cos x \ \text{при} \ \pi<x\le\frac{3}{2}\pi, \\ 0 \ \text{при} \ x>\frac{3}{2}\pi. \end{cases}$$

1) Определить вероятность попадания случайной величины X в интервал .
2) Найти математическое ожидание и дисперсию случайной величины X.

10. ***Исходные данные***: студенты некоторой группы, состоящей из 30 человек сдали экзамен по курсу «Информатика». Полученные студентами оценки образуют следующий ряд чисел:

4	4	3	3	2	5	2	3	3	4
3	4	4	2	5	2	3	3	4	4
3	3	4	4	2	5	5	2	3	3

Требуется:

1) составить дискретный вариационный ряд, при необходимости упорядочив его;

2) определить основные числовые характеристики ряда;

3) дать графическое представление ряда в виде полигона (гистограммы) распределения;

4) построить теоретическую кривую нормального распределения и проверить соответствие эмпирического и теоретического распределений по критерию Пирсона. При проверке статистической гипотезы о виде распределения принять уровень значимости а = 0,05

Литература

1. Вентцель Е.С. Теория случайных процессов и её инженерные приложения. М.: Кнорус, 2014. 448 с.
2. Гмурман В.Е. Теория вероятности и математическая статистика. М.: Юрайт, 2010. 479 с.: ил.
3. Гмурман В.Е. Руководство к решению задач по теории вероятности и математической статистике. М.: Юрайт, 2013. 404 с.
4. Горелова Г.В. Теория вероятности и математическая статистика в примерах и задачах с применением Excel. Ростов н/Д.: Феникс, 2006. 475 с.
5. Кочетков Е.С. Теория вероятностей и математическая статистика. М.: Инфра-М, 2005. 240 с.
6. Попов А.М. Теория вероятностей и математическая статистика. М.: Юрайт, 2011. 440 с.

Учебное издание

Рыжик Валентина Николаевна

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания для самостоятельной работы по дисциплине «Высшая математика»

Редактор Павлютина И.П.

Подписано в печать 11.05.2018 г. Формат 60х84 1/16.
Бумага офсетная. Усл.п.л. 4.59. Тираж 50 экз. Изд. №5944.

Издательство Брянского государственного Аграрного университета.
243365, Брянская область, Выгоничский район, с. Кокино, БГАУ