**Федеральное государственное образовательное бюджетное учреждение
высшего образования**

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»
Департамент Анализа данных, принятия решений и финансовых технологий**

А.В. Потемкин, И.И. Цыганок, М.Н. Фридман, И.М. Эйсымонт

АНАЛИЗ ДАННЫХ: ЧАСТЬ 2

Учебное пособие для студентов заочной формы обучения

Для бакалавров направлений
38.03.01 «Экономика» и 38.03.05 «Бизнес-информатика»

Москва 2019

Федеральное государственное образовательное бюджетное учреждение
высшего образования

«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»
Департамент Анализа данных, принятия решений и финансовых технологий

А.В. Потемкин, И.И. Цыганок, М.Н. Фридман, И.М. Эйсымонт

АНАЛИЗ ДАННЫХ: ЧАСТЬ 2

Учебное пособие для студентов заочной формы обучения

Для бакалавров направлений
38.03.01 «Экономика» и 38.03.05 «Бизнес-информатика»

Одобрено Департаментом анализа данных, принятия решений и финансовых технологий,
протокол №   2019 г.

Москва-2019

УДК 519.2(075.8)

ББК 22.17я73

П 64

Рецензент: И.Е.Денежкина, к.т.н., доцент департамента анализа данных, принятия решений и финансовых технологий.

А.В.Потемкин, М.Н. Фридман, И.И. Цыганок, И.М.Эйсымонт.

АНАЛИЗ ДАННЫХ: ЧАСТЬ 2. Учебное пособие для студентов заочной формы обучения для бакалавров направлений 38.03.01 «Экономика» и 38.03.05 «Бизнес-информатика» – М.: ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации», Департамента Анализа данных, принятия решений и финансовых технологий, 2019. - 118 с.

Дисциплина «Анализ данных» является базовой компонентой цикла математических и естественнонаучных дисциплин ФГОС ВО по направлениям 38.03.01 «Экономика» и 38.03.05 «Бизнес-информатика». Учебное пособие предназначено для студентов заочной формы обучения и содержит перечень изучаемых тем в соответствии с рабочей программой дисциплины «Анализ данных», решения типовых задач с подробными пояснениями, задачи для самостоятельного решения, варианты расчётно-аналитической работы, вопросы к экзамену, демонстрационные варианты билетов, таблицы, список основной и дополнительной литературы.

УДК 519.2(075.8)

ББК 22.17я73

Москва 2019

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

Студенты, обучающиеся по направлениям 38.03.01 «Экономика» и 38.03.05 «Бизнес-информатика», дисциплину «Анализ данных» изучают в течение двух семестров, учебным планом предусмотрено выполнение контрольной работы и расчётно-аналитической работы в качестве текущего контроля, промежуточная аттестация предполагает сдачу зачёта и экзамена.

Основное содержание дисциплины «Анализ данных» для бакалавров направлений 38.03.01 «Экономика» и 38.03.05 «Бизнес-информатика» и требования, предъявляемые образовательными стандартами к результатам освоения дисциплины, подробно изложены в рабочей программе. Ниже приводится перечень тем, изучаемых в рамках дисциплины.

Тема 1. Данные в экономике, их визуализация и предварительная обработка

1.1. Данные в экономике. Объекты, признаки и таблицы. Типы признаков в экономике и управлении: интервальные, порядковые, ранговые, дихотомические. Форматирование наборов данных как таблиц в Microsoft Excel. Гистограммы в Microsoft Excel. Условное форматирование в Microsoft Excel. Графики и диаграммы рассеяния в Microsoft Excel.

1.2. Инструменты описательной статистики в Microsoft Excel. Измерение центра распределения. Измерение разброса данных. Описательная статистика в надстройке «Анализ данных» Microsoft Excel. Диаграммы размаха в Microsoft Excel.

1.3. Визуализация качественных признаков в Microsoft Excel. Сводные таблицы и сводные диаграммы в Microsoft Excel. Таблицы сопряженности и парадокс Симпсона. Иерархия признаков в Microsoft Excel.

1.4. Предварительная обработка данных. Выбросы и их обработка в Microsoft Excel. Пропущенные значения и их обработка в Microsoft Excel.

Повторяющиеся строки и их обработка в Microsoft Excel. Синтетические признаки.

Тема 2. Случайные события

2.1. Основы комбинаторики. Правила суммы и произведения. Перестановки, размещения и сочетания без повторений. Перестановки, размещения и сочетания с повторениями. Формулы комбинаторики в Microsoft Excel.

2.2. Определение вероятности. Случайные события, их виды. Операции над событиями как операции над множествами. Классическая вероятностная схема. Схема геометрических вероятностей. Статистическая вероятность. Аксиоматическое построение теории вероятностей. Теорема сложения вероятностей. Обобщенная теорема сложения вероятностей.

2.3. Условные вероятности. Условная вероятность. Независимость событий. Формула полной вероятности. Формула Байеса. Функция СУММПРОИЗВ. Простейшие примеры применения теории вероятностей в экономике, управлении и финансах.

2.4. Последовательности испытаний. Биномиальная схема. Формула Бернулли. Формула Пуассона. Функции БИНОМ.РАСП и ПУАССОН.РАСП. Последовательности испытаний в экономике и управлении.

Тема 3. Случайные величины

3.1. Определение случайной величины. Понятие случайной величины. Функция распределения случайной величины. Свойства функции распределения. Индикатор события как простейшая случайная величина. Функция распределения индикатора события.

3.2. Дискретные случайные величины и их важнейшие числовые характеристики. Дискретная случайная величина. Ряд распределения и функция распределения дискретной случайной величины. Математическое ожидание, дисперсия и среднее квадратическое отклонение дискретной случайной величины.

3.3. Дискретные случайные величины, часто встречающиеся в экономической практике. Биномиальный закон распределения. Биномиальная модель ценообразования финансовых инструментов. Геометрический закон распределения. Закон распределения Пуассона. Простейший поток событий. Гипергеометрический закон распределения. Реализация моделей дискретных случайных величин в пакете Microsoft Excel при решении экономических задач. Сравнение случайных величин: отношение предпочтения, ожидаемая полезность, оптимальность по Парето.

3.4. Абсолютно непрерывные случайные величины и их важнейшие числовые характеристики. Абсолютно непрерывная случайная величина. Функция распределения и функция плотности распределения абсолютно непрерывной случайной величины. Свойства функции плотности распределения. Математическое ожидание, дисперсия и среднее квадратическое отклонение абсолютно непрерывной случайной величины.

3.5. Абсолютно непрерывные случайные величины, часто встречающиеся в экономической практике. Равномерный закон распределения. Показательный закон распределения. Нормальный закон распределения. Логарифмически нормальный закон распределения и ценообразование финансовых инструментов. Закон распределения Парето и задачи налогообложения. Законы распределения, важные в математической статистике (законы распределения Стьюдента, χ^2, Фишера — Снедекора). Реализация моделей абсолютно непрерывных случайных величин в пакете Microsoft Excel при решении экономических задач. Смеси распределений.

3.6. Моменты и критические границы случайной величины. Начальные и центральные моменты случайной величины. Асимметрия и эксцесс случайной величины. Квантили и процентные точки случайной величины. Вычисление квантилей и процентных точек в Microsoft Excel. Ценность под риском. Медиана и мода случайной величины.

3.7. Меры связи случайных величин. Случайные векторы и условные законы распределения. Условный ряд распределения (для дискретных случайных величин), условная плотность распределения (для непрерывных случайных величин). Условное математическое ожидание. Формула полного математического ожидания. Формула полной дисперсии. Ковариация и коэффициент корреляции. Портфель финансовых инструментов

3.8. Функции случайных величин. Функции одной случайной величины. Функции нескольких случайных величин. Формула композиции. Композиция равномерных случайных величин.

Тема 4. Предельные теоремы теории вероятностей

4.1. Закон больших чисел. Массовые случайные явления в экономике. Теорема Чебышёва и оценка математического ожидания. Теорема Бернулли и оценка вероятности. Обсуждение условий статистической устойчивости.

4.2. Центральная предельная теорема. Теорема Леви. Интегральная теорема Муавра — Лапласа. Математические основы теории страхования. Метод Монте-Карло. Моделирование случайных величин в Microsoft Excel. Функция СЛЧИС и программа «Генерация случайных чисел» надстройки «Анализ данных» пакета Microsoft Excel. Место центральной предельной теоремы в изучении статистических закономерностей в экономике, финансах и управлении.

Тема 5. Оценка параметров

5.1. Основы выборочного метода. Предмет и задачи математической статистики. Генеральная и выборочная совокупности. Случайная и конкретная выборки. Случайная повторная и случайная бесповторная выборка. Соотношение между предельной ошибкой выборки, уровнем значимости (риском) и объемом выборки. Использование этого соотношения в организации выборочных обследований.

5.2. Оценка плотности распределения и функции распределения. Вариационный ряд. Выборочная случайная величина (статистический ряд распределения). Интервальный вариационный ряд. Полигон частот, кумулята.

Оценка числовых характеристик генеральной случайной величины с помощью выборочной случайной величины. Выборочное среднее как оценка математического ожидания. Относительная частота как оценка вероятности. Выборочная дисперсия как оценка дисперсии. Программа «Гистограмма» надстройки «Анализ данных» пакета Microsoft Excel.

5.3. Точечные оценки параметров. Понятие точечной оценки параметра генеральной совокупности. Свойства точечных оценок: состоятельность, несмещенность, эффективность. Выборочное среднее как состоятельная, несмещенная и эффективная оценка математического ожидания генеральной случайной величины. Смещенность выборочной дисперсии как оценки дисперсии генеральной случайной величины. Исправленная выборочная дисперсия как несмещенная и состоятельная оценка дисперсии генеральной случайной величины. Методы построения точечных оценок: метод моментов, метод максимального правдоподобия. Построение оценок параметров распределений случайных величин, применяемых в экономике и управлении.

5.4. Интервальные оценки параметров. Понятие интервальной оценки параметра генеральной совокупности. Точные интервальные оценки вероятности, математического ожидания, дисперсии и коэффициента корреляции. Поправка на конечный объем генеральной совокупности. Асимптотический подход к интервальному оцениванию.

Тема 6. Проверка статистических гипотез

6.1. Статистические гипотезы. Понятие статистической гипотезы. Виды статистических гипотез: параметрические и непараметрические, простые и сложные. Критерий проверки гипотезы, критическое множество. Проверка гипотез с помощью интервальных оценок. Ошибки первого и второго родов. Мощность критерия. Наиболее мощный критерий.

6.2. Параметрические критерии. Проверка гипотезы о равенстве математического ожидания теоретическому значению. Проверка гипотезы о равенстве двух математических ожиданий. Проверка гипотезы о равенстве дисперсии теоретическому значению. Проверка гипотезы о равенстве двух дисперсий. Проверка гипотезы о равенстве вероятности события теоретическому значению. Проверка гипотезы о равенстве двух вероятностей. Проверка гипотез о значимости коэффициента корреляции. Использование аппарата проверки гипотез в экономике и управлении. Реализация критериев проверки статистических гипотез в пакете Microsoft Excel.

6.3. Критерии согласия. Критерий согласия χ^2 Пирсона. Критерий χ^2 Пирсона при неизвестных параметрах распределения.

Тема 7. Дисперсионный анализ

7.1. Однофакторный дисперсионный анализ. Понятие о дисперсионном анализе. Задача дисперсионного анализа и классификация его моделей. Однофакторная детерминированная модель дисперсионного анализа: проверяемые гипотезы, выборочное дисперсионное тождество, дисперсионная таблица и проверка гипотез, выборочные коэффициенты детерминации, оценка параметров модели и проверка гипотез. Однофакторная случайная модель дисперсионного анализа: проверяемые гипотезы, выборочное дисперсионное тождество, дисперсионная таблица и проверка гипотез, выборочные коэффициенты детерминации, оценка параметров модели и проверка гипотез. Реализация моделей однофакторного дисперсионного анализа в пакете Microsoft Excel. Примеры экономических и социальных задач, решаемых с помощью однофакторного дисперсионного анализа.

7.2. Двухфакторный дисперсионный анализ. Двухфакторная детерминированная модель дисперсионного анализа с одним и более наблюдением в клетке: проверяемые гипотезы, выборочное дисперсионное тождество, дисперсионная таблица и проверка гипотез, выборочные коэффициенты детерминации, оценка параметров модели и проверка гипотез. Двухфакторная случайная модель дисперсионного анализа с одним и более наблюдением в клетке: проверяемые гипотезы, выборочное дисперсионное тождество, дисперсионная таблица и проверка гипотез, выборочные коэффициенты детерминации, оценка параметров модели и проверка гипотез. Двухфакторная смешанная модель дисперсионного анализа с одним и более наблюдением в клетке: проверяемые гипотезы, выборочное дисперсионное тождество, дисперсионная таблица и проверка гипотез, выборочные коэффициенты детерминации, оценка параметров модели и проверка гипотез. Реализация моделей двухфакторного дисперсионного анализа в пакете Microsoft Excel. Примеры экономических и социальных задач, решаемых с помощью двухфакторного дисперсионного анализа.

Тема 8. Основы непараметрической статистики

8.1. Таблицы сопряженности. Критерий χ^2 для проверки независимости компонент случайной величины. Критерий χ^2 для проверки однородности данных.

8.2. Непараметрические критерии. Проверка гипотез на малых выборках. Критерий знаков. Распределение Вилкоксона и его критические границы. Непараметрическая точечная оценка математического ожидания. Непараметрическая интервальная оценка математического ожидания. Критерий Вилкоксона (парный критерий знаковых рангов). Распределение Вилкоксона — Манна — Уитни и его критические границы. Непараметрическая точечная оценка теоретической величины сдвига. Непараметрическая интервальная оценка теоретической величины сдвига. Критерий Вилкоксона — Манна — Уитни (непараметрический критерий сравнения математических ожиданий для независимых выборок). Примеры применения непараметрических критериев в экономике.

8.3. Ранговая корреляция. Коэффициент ранговой корреляции Спирмена. Коэффициент ранговой корреляции Кендалла. Коэффициент конкордации. Проверка гипотез о значимости ранговых коэффициентов корреляции. Примеры использования ранговой корреляции в экономике.

Тема 9. Основы машинного обучения

9.1. Задачи машинного обучения. Обучение с учителем и обучение без учителя. Классы задач машинного обучения: регрессия, классификация, кластерный анализ, поиск аномалий. Примеры задач машинного обучения в экономике, управлении и финансах.

9.2. Линейная регрессия. Постановка задачи регрессионного анализа. Парная линейная регрессия. Множественная линейная регрессия. Точечный и интервальный прогноз по модели регрессии. Примеры задач регрессии в экономике. Понятие о гетероскедастичности и автокорреляции.

9.3. Классификация с обучением. Постановка задачи классификации с обучением. Логистическая регрессия. Понятие о деревьях решений. Кредитный скоринг.

9.4. Кластерный анализ и поиск аномалий. Постановка задачи кластерного анализа. Метод К-средних. Сегментирование потребителей. Понятие о методах машинного обучения в задачах поиска аномалий.

В предлагаемом учебном пособии содержится необходимый методический материал, позволяющий студентам выполнить расчётно-аналитическую работу по темам 2, 3, 4, 5, 6 и 9. Для подготовки к экзамену по всему курсу дисциплины необходимо будет использовать материал, изложенный и в учебном пособии «Анализ данных: Часть 1».

§ 1. ПОСЛЕДОВАТЕЛЬНОСТИ ИСПЫТАНИЙ. БИНОМИАЛЬНАЯ СХЕМА.

Пример 1. Студент приобрел пять лотерейных билетов. Вероятность выигрыша по одному билету составляет 0,2. Найти вероятность того, что студент выиграет:

а) по трем лотерейным билетам;

б) не менее чем по трем билетам;

в) хотя бы по одному билету.

Определить наивероятнейшее число выигрышных билетов.

Решение. Испытание состоит в проверке билета на выигрыш. По условию число таких испытаний составляет $n=5$. В каждом испытании наступает или не наступает событие *A – проверяемый билет содержит выигрыш*. Очевидно, что наступление события *A* в каждом предыдущем испытании не изменяет его вероятности в последующих и, следовательно, испытания являются независимыми. При этом, вероятность наступления события *A* в каждом испытании постоянна и равна $p=0{,}2$, и, таким образом, вероятность того, что событие *A* не наступит, равна $q=1-p=1-0{,}2=0{,}8$. На основании *вышесказанного* можно сделать вывод о том, что мы имеем дело с повторными независимыми испытаниями, и при решении задачи можно использовать формулу Бернулли.

а) Пусть событие *B – студент выиграл по трем лотерейным билетам*. Следовательно, число испытаний, в которых ожидается наступление события *A,* равно $m=3$. Искомую вероятность найдем по формуле Бернулли:

$$P(B)=P_{3,5}=P_5(3)=C_5^3p^3q^{5-3}=\frac{5!}{3!\cdot(5-3)!}\cdot 0{,}2^3\cdot 0{,}8^2=0{,}0512.$$

б) Пусть событие *C – студент выиграл не менее чем по трем лотерейным билетам*. Такому событию благоприятствуют три случая:

событие *A* наступает три, четыре или пять раз. Все эти события несовместны, поэтому, согласно теореме сложения вероятностей несовместных событий, вероятность суммы несовместных событий будет равна сумме вероятностей каждого события, включенного в сумму, т.е. имеем:

$$P(C) = P_{3,5} + P_{4,5} + P_{5,5} = C_5^3 p^3 q^2 + C_5^4 p^4 q^1 + C_5^5 p^5 q^0 =$$
$$= 10 \cdot 0{,}2^3 \cdot 0{,}8^2 + 5 \cdot 0{,}2^4 \cdot 0{,}8 + 0{,}2^5 = 0{,}05792.$$

в) Пусть событие *D* – *студент выиграл хотя бы по одному лотерейному билету*. Выиграть хотя бы по одному билету означает выигрыш либо по одному, либо по двум, либо по трем, либо по четырем или, наконец, по пяти билетам, т.е. событие *D* представляет сумму всех этих несовместных событий. Поэтому, здесь также можно было бы воспользоваться теоремой сложения вероятностей несовместных событий. Но достаточно большое число слагаемых делает расчет весьма громоздким, и чтобы избежать этого, проще перейти к противоположному событию $\bar{D}$ – *все билеты без выигрышей*. Для такого события число испытаний, в которых ожидается наступление события *A*, равно $m = 0$. На основании свойства вероятностей противоположных событий, вероятность искомого события будет равна:

$$P(D) = 1 - P(\bar{D}) = 1 - P_{0,5} = 1 - C_5^0 p^0 q^5 = 1 - 0{,}8^5 = 0{,}67232 \ .$$

Наиболее вероятное число успехов m_0 удовлетворяет следующему двойному неравенству:

$$np - q \le m_0 \le np + p.$$

Если $np + p$ – целое число, то наивероятнейших чисел два: $m_{0,1} = np - q$ и $m_{0,2} = np + p$. Если $np + p$ – не целое число, то наивероятнейшее число равно $m_0 = [np + p]$, где символ $[x]$ обозначает целую часть числа *x*.

В нашем случае наивероятнейшее число выигрышных билетов будет равно:

$$m_0 = [np + p] = [5 \cdot 0{,}2 + 0{,}2] = [1{,}2] = 1.$$

Пример 2. В среднем по 15% договоров страховая компания выплачивает страховую сумму. Найти вероятность того, что из 200 договоров с выплатой страховой суммы будет связано:

а) 20 договоров;

б) наивероятнейшее число договоров;

в) от 25 до 45 договоров включительно;

г) не более 40 договоров;

д) от 25 до 35 договоров.

Решение. Очевидно, что здесь имеют место повторные независимые испытания. Каждое испытание связано с наступлением или не наступлением страхового случая. Пусть событие A состоит в том, что страховой случай наступил. По условию задачи вероятность такого события равна $p = P(A) = 0{,}15$ и, соответственно, вероятность не наступления будет равна $q = P(\bar{A}) = 1 - p = 1 - P(A) = 1 - 0{,}15 = 0{,}85$. Так как число испытаний $n = 200$ достаточно велико, применять здесь формулу Бернулли нецелесообразно, и надо воспользоваться асимптотическими формулами. В частности, выполнены все условия применимости локальной и интегральной теорем Муавра-Лапласа, ибо число испытаний $n = 200$ достаточно велико и $npq = 200 \cdot 0{,}15 \cdot 0{,}85 = 25{,}5 > 20$.

а) Применим локальную теорему Муавра-Лапласа. По условию задачи $m = 20$. Определяем соответствующее ему значение $x = \dfrac{20 - 200 \cdot 0{,}15}{\sqrt{200 \cdot 0{,}15 \cdot 0{,}85}} \approx -1{,}98$, а затем находим вероятность искомого события:

$$P_{20,200} \approx \frac{1}{\sqrt{200 \cdot 0{,}15 \cdot 0{,}85}} \cdot f(-1{,}98) = \frac{1}{\sqrt{25{,}5}} \cdot f(1{,}98) = \frac{0{,}0562}{\sqrt{25{,}5}} \approx 0{,}0111.$$

При вычислении воспользовались таблицей значений функции Гаусса (приложение 1) и свойством четности функции Гаусса.

б) Сначала найдем наивероятнейшее число страховых случаев по формуле $m_0 = [np + p]$, согласно которой $m_0 = [200 \cdot 0{,}15 + 0{,}15] = 30$. Для этого значения находим:

$$x = \frac{30 - 200 \cdot 0{,}15}{\sqrt{200 \cdot 0{,}15 \cdot 0{,}85}} = 0 \qquad \text{и}$$

$$P_{30,200} \approx \frac{1}{\sqrt{200 \cdot 0{,}15 \cdot 0{,}85}} \cdot f(0) = \frac{0{,}3989}{\sqrt{25{,}5}} \approx 0{,}07899.$$

в) Воспользуемся интегральной теоремой Муавра-Лапласа для вычисления $P_{200}(25 \le m \le 45)$, предварительно определив аргументы функции Лапласа:

$$x_1 = \frac{25 - 200 \cdot 0{,}15}{\sqrt{200 \cdot 0{,}15 \cdot 0{,}85}} \approx -0{,}99 \quad \text{и} \qquad x_2 = \frac{45 - 200 \cdot 0{,}15}{\sqrt{200 \cdot 0{,}15 \cdot 0{,}85}} \approx 2{,}97.$$

Далее

$$P_{200}(25 \le m \le 45) \approx \frac{1}{2}\left[\Phi(2{,}97) - \Phi(-0{,}99)\right] = \frac{1}{2}\left[\Phi(2{,}97) + \Phi(0{,}99)\right] =$$
$$= \frac{1}{2}(0{,}9970 + 0{,}6778) = 0{,}8374.$$

При вычислении воспользовались нечетностью функции Лапласа и таблицей удвоенных значений функции Лапласа (приложение 2-2).

в) Необходимо найти $P_{200}(m \le 40)$. Так как число договоров не может быть отрицательным, то определение последней вероятности эквивалентно вычислению вероятности $P_{200}(0 \le m \le 40)$. Определив

$$x_1 = \frac{0 - 200 \cdot 0{,}15}{\sqrt{200 \cdot 0{,}15 \cdot 0{,}85}} \approx -5{,}94 \text{ и } x_2 = \frac{40 - 200 \cdot 0{,}15}{\sqrt{200 \cdot 0{,}15 \cdot 0{,}85}} \approx 1{,}98,$$

получим:

$$P_{200}(0 \le m \le 40) \approx \frac{1}{2}\left[\Phi(1{,}98) - \Phi(-5{,}94)\right] = \frac{1}{2}\left[\Phi(2{,}97) + \Phi(5{,}94)\right] \approx$$
$$\approx \frac{1}{2}(0{,}9523 + 1) = 0{,}97615.$$

При вычислении здесь воспользовались нечетностью функции Лапласа, таблицей удвоенных значений функции Лапласа, а также тем, что если аргумент удвоенной функции Лапласа больше четырех, то ее значение можно принимать равным единице.

д) Вероятность $P_{200}\left(25 \le m \le 35\right)$ можно было найти так же, как и в предыдущем случае, но проще это сделать, используя следствие из интегральной теоремы Муавра-Лапласа.

Заметим, что границы промежутка 25 и 35 симметричны относительно значения $np = 200 \cdot 0{,}15 = 30$. Поэтому,

$$P_{200}\left(25 \le m \le 35\right) = P_{200}\left(\left|m-30\right| \le 5\right) \approx \Phi\left(\frac{5}{\sqrt{200 \cdot 0{,}15 \cdot 0{,}85}}\right) = \Phi\left(0{,}99\right) = 0{,}6778.$$

Пример 3. По статистическим данным известно, что в некоторой местности в среднем на каждые 100 семей приходится 30 автомобилей.

1. Найти вероятность того, что из 1000 семей доля имеющих автомобиль:

а) будет заключена в пределах от 0,28 до 0,33;

б) будет отличаться от вероятности этого события не более чем на 0,02 (по абсолютной величине).

2. При каком числе семей можно утверждать с надежностью, не меньшей 0,9545, что доля семей, имеющих автомобиль, будет заключена в границах от 0,28 до 0,32?

Решение. Пусть событие A состоит в том, что некоторая семья имеет автомобиль. На основании статистического определения вероятности, можно сказать, что вероятность такого события будет равна $p = P\left(A\right) = \frac{30}{100} = 0{,}3$.

1. Так как число испытаний $n = 1000$ достаточно велико и условие $npq = 1000 \cdot 0{,}3 \cdot 0{,}7 = 210 > 20$ выполнено, то здесь можно применить следствие из интегральной теоремы Муавра-Лапласа

а) Необходимо найти $P_{1000}\left(0,28 \le \frac{m}{n} \le 0,33\right)$. Для этого вначале определим аргументы функции Лапласа:

$$z_1 = \frac{0,28-0,3}{\sqrt{\frac{0,3\cdot 0,7}{1000}}} = -1,38 \text{ и } z_2 = \frac{0,33-0,3}{\sqrt{\frac{0,3\cdot 0,7}{1000}}} = 2,07\,.$$

Тогда:

$$P_{1000}\left(0,28 \le \frac{m}{n} \le 0,33\right) \approx \frac{1}{2}\left[\Phi(2,07)-\Phi(-1,38)\right] = \frac{1}{2}\left[\Phi(2,07)+\Phi(1,38)\right] =$$

$$= \frac{1}{2}(0,9616+0,8324) = 0,8970.$$

б) По следствию из интегральной теоремы Муавра-Лапласа для доли найдем:

$$P_{1000}\left(\left|\frac{m}{n}-0,3\right| \le 0,02\right) \approx \Phi\left(\frac{0,02}{\sqrt{\frac{0,3\cdot 0,7}{1000}}}\right) = \Phi(1,38) = 0,8324\,.$$

2. По условию

$$P_{1000}\left(0,28 \le \frac{m}{n} \le 0,32\right) \ge 0,9545 \text{ или } P_{1000}\left(\left|\frac{m}{n}-0,3\right| \le 0,02\right) \ge 0,9545\,.$$

Заменяя левую часть неравенства согласно следствию из интегральной теоремы Муавра-Лапласа для доли функцией Лапласа, приходим к следующему соотношению:

$$\Phi\left(\frac{0,02}{\sqrt{\frac{0,3\cdot 0,7}{n}}}\right) \ge 0,9545\,.$$

Воспользовавшись таблицей удвоенных значений функции Лапласа, определяем ее аргумент, соответствующий значению функции, равному 0,9545. По таблице получаем, что аргумент равен 2, т.е.:

$$\Phi\left(\frac{0{,}02}{\sqrt{\frac{0{,}3\cdot 0{,}7}{n}}}\right)\ge \Phi(2).$$

Далее, используя свойство возрастания функции Лапласа, получим:

$$\frac{0{,}02}{\sqrt{\frac{0{,}3\cdot 0{,}7}{n}}}\ge 2.$$

Разрешая последнее соотношение относительно n, найдем, что:

$$\sqrt{n}\ge \frac{\sqrt{0{,}3\cdot 0{,}7}}{0{,}01} \text{ или } n\ge \frac{0{,}3\cdot 0{,}7}{0{,}01^2}=2100.$$

Таким образом, число обследуемых семей должно быть не менее 2100, т.е. увеличено более чем в два раза.

Пример 4. Прядильщица обслуживает 1000 веретен. Вероятность обрыва нити на одном веретене в течение одной минуты равна 0,004. Найти вероятность того, что в течение одной минуты обрыв произойдет:

а) на пяти веретенах;

б) хотя бы на двух веретенах.

Решение. В данном случае испытание состоит в проверке веретен на обрыв нити. По условию проводится достаточно большое число $n=1000$ испытаний, в каждом из которых вероятность наступления события A – обрыв нити на одном веретене – одна и та же и равна $p=P(A)=0{,}004$. Вероятность такого события достаточно мала и число $\lambda=np=1000\cdot 0{,}004=4<10$. Из сказанного выше вытекает, что выполнены все условия применимости формулы Пуассона.

а) Подставляя значения параметров $\lambda=4$ и $m=5$, найдем вероятность того, что в течение одной минуты обрыв произойдет только на пяти веретенах: $P_{5,1000}\approx\frac{4^5 e^{-4}}{5!}$. Числовое значение этой вероятности при указанных выше параметрах $\lambda = 4$ и $m = 5$ найдем по таблице значений функции Пуассона (приложение 3):

$$P_{5,1000} \approx 0{,}1563.$$

б) Вероятность того, что обрыв произойдет хотя бы на двух веретенах, можно вычислить как сумму вероятностей большого числа несовместных событий:

$$P_{1000}(m \geq 2) = P_{2,1000} + P_{3,1000} + \ldots + P_{1000,1000}.$$

Однако, очевидно, что такого типа вероятности проще вычислять через вероятность противоположного события, состоящего в том, что обрыва нити либо не будет ($m = 0$), либо обрыв произойдет только на одном веретене ($m = 1$).

Используя формулу для вероятностей противоположных событий и теорему сложения вероятностей несовместных событий, получим:

$$P_{1000}(m \geq 2) = 1 - P_{1000}(m < 2) = 1 - \left(P_{1000}(m = 0) + P_{1000}(m = 1)\right) \approx$$
$$\approx 1 - \left(0{,}0183 + 0{,}0733\right) = 0{,}9084.$$

Замечание. Последнюю вероятность нельзя вычислить по интегральной теореме Муавра-Лапласа, так как не выполнено одно из условий ее применимости, а именно $npq = 1000 \cdot 0{,}004 \cdot 0{,}996 = 3{,}984 < 20$.

§ 2. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ ПО ТЕМЕ «ПОСЛЕДОВАТЕЛЬНОСТИ ИСПЫТАНИЙ. БИНОМИАЛЬНАЯ СХЕМА»

1. По статистическим данным, реклама новых банковских проектов привлекает 15 % населения, прочитавшего ее. Если 200 человек прочитали рекламу, то какова вероятность того, что:

а) ровно 50 человек воспользуются ею;

б) число участников превысит 40;

в) число участников будет заключаться от 25 до 100;

г) число участников будет заключаться от 25 до 45.

Найти наивероятнейшее число участников и его вероятность.

2. По статистическим данным, в 20% случаев коммерческому банку удается привлечь имеющиеся у населения сбережения на новых, более выгодных условиях. Найти вероятность того, что среди населения данного округа численностью 1500 человек доля граждан, желающих вложить свои сбережения в коммерческий банк на новых условиях:

а) будет заключена в пределах от 0,28 до 0,33;

б) будет отличаться от вероятности этого события не более чем на 0,03 (по абсолютной величине).

При какой численности населения можно утверждать, что с вероятностью 0,9545 доля граждан, желающих вложить свои сбережения в коммерческий банк на новых условиях, будет заключена в границах от 0,25 до 0,35?

3. Фирма, занимающаяся реализацией оргтехники, рассылает рекламные проспекты по организациям. По статистике, примерно в одном случае из 10 при этом следует заказ. Найти вероятность того, что среди 600 организаций, получивших рекламные проспекты, доля организаций, заказавших оргтехнику:

а) заключена в пределах от 0, 08 до 0, 15;

б) будет отличаться от вероятности этого события не более чем на 0,01 по абсолютной величине).

Сколько рекламных проспектов следует разослать, чтобы с вероятностью 0,97 можно было ожидать, то доля заказов будет заключена в границах от 0,08 до 0,12?

4. При отливке блока цилиндров двигателя с вероятностью 0,002 возникает брак. Найти вероятность того, что среди 500 отливок:

а) будут 3 бракованные;

б) будет не менее 3 бракованных.

5. Вероятность того, что смартфон определенного производителя выйдет из строя в течение гарантийного срока, равна 0,04. Найти вероятность того, что среди 200 смартфонов этого производителя выйдут из строя в течение гарантийного срока:

а) 3 смартфона;

б) хотя бы один из них.

§ 3. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Пример 5. Пусть плотность распределения непрерывной случайной величины ξ имеет вид:

$$\varphi(x) = \begin{cases} \dfrac{1}{2} & \text{при } x \in [0;2], \\ 0 & \text{в остальных случаях.} \end{cases}$$

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

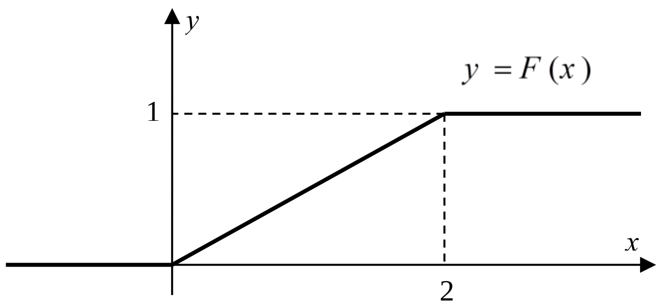

Рис. 1

Решение. Пусть $x < 0$. Тогда

$$F(x) = \int_{-\infty}^{x} \varphi(t)dt = \int_{-\infty}^{x} 0\,dt = 0.$$

Если $x \in [0;2]$, то

$$F(x) = \int_{-\infty}^{x} \varphi(t)\,dt = \int_{-\infty}^{0} 0dt + \int_{0}^{x} \frac{1}{2}dt = 0 + \frac{1}{2}t\Big|_{0}^{x} = \frac{1}{2}x.$$

Если $x > 2$, то

$$F(x)=\int_{-\infty}^{x}\varphi(t)\,dt=\int_{-\infty}^{0}0dt+\int_{0}^{2}\frac{1}{2}dt+\int_{2}^{x}0\,dt=0+\frac{1}{2}t\Big|_{0}^{2}+0=\frac{1}{2}(2-0)=1.$$

Таким образом, окончательно, искомая функция распределения имеет вид

$$F(x)=\begin{cases}0 & \text{при } x<0,\\ \frac{1}{2}x & \text{при } x\in[0;\ 2],\\ 1 & \text{при } x>2.\end{cases}$$

График полученной функции распределения представлен на рис. 5.

Вычислим математическое ожидание и дисперсию. Согласно определению математического ожидания и дисперсии, имеем

$$M\xi=\int_{-\infty}^{+\infty}x\varphi(x)dx=\int_{-\infty}^{0}x\cdot 0\,dx+\int_{0}^{2}x\cdot\frac{1}{2}dx+\int_{2}^{+\infty}x\cdot 0\,dx=\frac{1}{2}\cdot\frac{x^2}{2}\Big|_{0}^{2}=1.$$

$$M\xi^2=\int_{-\infty}^{+\infty}x^2\varphi(x)dx=\int_{-\infty}^{0}x^2\cdot 0\,dx+\int_{0}^{2}x^2\cdot\frac{1}{2}dx+\int_{2}^{+\infty}x^2\cdot 0\,dx=\frac{1}{2}\cdot\frac{x^3}{3}\Big|_{0}^{2}=\frac{4}{3},$$

$$D\xi=M\xi^2-M^2\xi=\frac{4}{3}-1^2=\frac{1}{3}.$$

Пример 6. При каком значении параметра α функция

$$F(x)=\begin{cases}0 & \text{при } x\le 0,\\ \alpha x^2 & \text{при } 0<x\le 4,\\ 1 & \text{при } x>4.\end{cases}$$

будет функцией распределения непрерывной случайной величины.

Вычислить ее математическое ожидание, дисперсию и среднее квадратическое отклонение.

Решение. Значение параметра α можно найти несколькими способами.

Первый способ. Воспользуемся свойством непрерывности функции $F(x)$. В частности, функция непрерывна в точке $x=4$. Поэтому, предел функции при x стремящемся к 4 слева должен быть равен пределу функции при x стремящемся к 4 справа, то есть

$$\lim_{x\to 4-0}F(x)=\lim_{x\to 4-0}\alpha x^2=4^2\alpha=16\alpha\ ,\ \lim_{x\to 4+0}F(x)=\lim_{x\to 4+0}1=1.$$

Из равенства пределов вытекает, что $16\alpha = 1$. Откуда $\alpha = \dfrac{1}{16}$.

Второй способ. Определим параметр распределения α, используя функцию плотности. Функция плотности распределения представляет собой первую производную от функции распределения:

$$\varphi(x) = F'(x) = \begin{cases} 0 & \text{при } x \le 0, \\ 2\alpha x & \text{при } 0 < x \le 4, \\ 0 & \text{при } x > 4. \end{cases}$$

Вычислим параметр α из условия нормировки функции плотности – несобственный интеграл от функции плотности на всей числовой оси должен быть равен единице. Так как рассматриваемая функция плотности отлична от нуля только на конечном промежутке, то несобственный интеграл от нее преобразуется к определенному интегралу на конечном промежутке $[0; 4]$. Следовательно, $\int\limits_0^4 2\alpha x dx = 1$. Вычисляя последний интеграл, имеем:

$$2\alpha \frac{x^2}{2}\Bigg|_0^4 = \alpha\left(4^2 - 0^2\right) = 16\alpha = 1.$$

Разрешая последнее соотношение относительно параметра α, получим:

$$\alpha = \frac{1}{16}.$$

Определим числовые характеристики данной случайной величины. Случайная величина ξ распределена на конечном промежутке (имеет функцию плотности, отличную от нуля на конечном промежутке), то согласно определению математического ожидания непрерывной случайной величины, получим:

$$M\xi = \int\limits_a^b x\varphi(x)dx = \int\limits_0^3 x\frac{2}{9}xdx = \frac{2}{9}\frac{x^3}{3}\Bigg|_0^3 = \frac{2}{9\cdot 3}\left(3^3 - 0^3\right) = 2.$$

Для вычисления дисперсии, сначала найдем математическое ожидание квадрата случайной величины:

$$M\xi^2=\int_a^b x^2\varphi(x)\,dx=\int_0^3 x^2\frac{2}{9}x\,dx=\frac{2}{9}\frac{x^4}{4}\Bigg|_0^3=\frac{1}{18}\left(3^4-0^4\right)=\frac{9}{2}$$

Тогда по свойству дисперсии, получим:

$$D\xi=M\xi^2-M^2\xi=\frac{9}{2}-2^2=\frac{1}{2}=0{,}5.$$

Среднее квадратическое отклонение равно:

$$\sigma_\xi=\sqrt{D\xi}=\sqrt{0{,}5}\approx 0{,}707.$$

Пример 7. Коробки с конфетами упаковываются автоматически. Их средний вес равен 300 г. Известно, что вес коробок с конфетами имеет нормальное распределение. При этом 5% коробок имеют вес, меньший 290 г. Записать функцию плотности распределения. Каков процент коробок, вес которых более 305 г?

Решение. Пусть случайная величина ξ – вес коробки конфет. По условию задачи случайная величина ξ имеет нормальное распределение. Найдем параметры этого распределения. Математическое ожидание случайной величины – это средний вес коробки, т.е. $M\xi=300$. Второй параметр распределения – среднее квадратическое отклонение случайной величины ξ – найдем из условия, что 5% коробок с конфетами имеют вес, меньший 290 г. Последнее означает, что вероятность того, что случайная величина ξ примет значение, меньшее 290, будет равна 0,05, т.е. $P(\xi<290)=0{,}05$.

Если случайная величина распределена по нормальному закону, то ее функцию распределения можно представить через функцию Лапласа:

$$F(x)=\frac{1}{2}+\frac{1}{2}\Phi\left(\frac{x-a}{\sigma}\right).$$

Далее, используя определения функции распределения случайной величины и свойства нечетности функции Лапласа $\Phi(-x) = -\Phi(x)$, вероятность $P(\xi < 290)$ можно записать в виде:

$$P(\xi < 290) = F(290) = \frac{1}{2} + \frac{1}{2}\Phi\left(\frac{290-300}{\sigma}\right) = \frac{1}{2} - \frac{1}{2}\Phi\left(\frac{10}{\sigma}\right) = 0{,}05.$$

Разрешая последнее соотношение относительно функции Лапласа, получим:

$$\Phi\left(\frac{10}{\sigma}\right) = 0{,}9.$$

Используя таблицу удвоенных значений функции Лапласа (приложение 2-2), найдем аргумент функции Лапласа:

$$\frac{10}{\sigma} = 1{,}96.$$

Откуда, $\sigma = \frac{10}{1{,}96} \approx 5{,}1$.

Таким образом, функция плотности распределения данной случайной величины имеет следующий вид:

$$\varphi(x) = \frac{1}{5{,}1\sqrt{2\pi}} e^{-\frac{(x-300)^2}{2\cdot 5{,}1^2}} = \frac{1}{5{,}1\sqrt{2\pi}} e^{-\frac{(x-300)^2}{52{,}02}}.$$

Зная параметры распределения, можно найти вероятность того, что вес коробки конфет будет составлять более 305 г.:

$$P(\xi > 305) = 1 - P(\xi \le 305) = 1 - F(305) =$$

$$= 1 - \left(\frac{1}{2} + \frac{1}{2}\Phi\left(\frac{305-300}{5{,}1}\right)\right) =$$

$$= \frac{1}{2} - \frac{1}{2}\Phi(0{,}98) = 0{,}5 - 0{,}5 \cdot 0{,}6729 = 0{,}16355.$$

§ 4. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ ПО ТЕМЕ «СЛУЧАЙНЫЕ ВЕЛИЧИНЫ»

6. Функция распределения непрерывной случайной величины ξ имеет вид:

$$F(x) = \begin{cases} 0, & \text{если } x \le 0, \\ \frac{1}{36}x^2, & \text{если } 0 < x \le 6, \\ 1, & \text{если } x > 6. \end{cases}$$

Найти:

а) математическое ожидание $M(\xi)$;

б) дисперсию $D(\xi)$;

в) $P(0{,}6 < \xi < 7)$.

7. Плотность вероятности случайной величины равна

$$\phi(x) = \begin{cases} -\frac{x^3}{4} & \text{при } -2 \le x \le 0, \\ 0 & \text{при всех остальных } x. \end{cases}$$

Найти функцию распределения этой случайной величины, ее математическое ожидание, и вероятность того, что она примет значения из промежутка (–1; 1).

8. При каком значении параметра a функция

$$\phi(x) = \begin{cases} 0{,}2 & \text{при } 1 \le x \le a, \\ 0 & \text{при всех остальных } x. \end{cases}$$

является плотностью вероятности некоторой случайной величины?

Найти ее функцию распределения, математическое ожидание и дисперсию, а также вероятность принять значения из промежутка (0;3).

9. Случайная величина ξ нормально распределена. Известно, что $M\xi = -2$, $D\xi = 1$.

Найти:

а) параметры a и σ^2закона распределения;

б) плотность вероятности случайной величины ξ и ее значения в точках $x = -1$, $x = 0$, $x = 2$;

в) вероятности $P(-2 < \xi < 0)$ и $P(\xi > 1)$.

10. Известно, что время службы электролампочек имеет нормальное распределение со средним значением 600 час. При этом 5% лампочек имеют срок службы, меньший 550час. Записать функцию плотности распределения. Каков процент лампочек со сроком службы более 620 часов?

11. Месячный расход воды в населенном пункте является случайной величиной, распределенной по нормальному закону. Найти его математическое ожидание и дисперсию и написать выражение плотности вероятности этого закона, если известно, что $P(X < 300) = 0{,}5$ и $P(250 \leq X \leq 350) = 0{,}9876$.

§ 5. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

Пример 8. Отделение банка обслуживает в среднем 100 клиентов в день. Оценить вероятность того, что сегодня в отделении банка будет обслужено:

а) более 150 клиентов;

б) не более 200 клиентов.

Решение. Пусть случайная величина ξ — число клиентов отделения банка, обслуживаемых за день. Очевидно, что ξ может принимать только неотрицательные значения и по условию $M\xi = 100$. Следовательно, а) необходимо оценить вероятность того, что случайная величина □ примет значение более 150.

Согласно лемме Чебышева, получим:

$$P(\xi > 150) \leq \frac{100}{150} = \frac{2}{3} \approx 0{,}67.$$

б) после перехода к противоположному событию определяем, что

$$P(\xi \leq 200) = 1 - P(\xi > 200) \geq 1 - \frac{100}{200} = 0{,}5.$$

Пример 9. Средняя температура в квартирах многоэтажного дома в период отопительного сезона равна 20°С, а среднее квадратическое отклонение равно 2°С.

1) Оценить с помощью неравенства Чебышева вероятность того, что температура в квартире отклонится от средней по абсолютной величине не более чем на 5 °С.

2) Найти эту вероятность, предполагая, что температура в квартире есть случайная величина, распределенная по нормальному закону.

Решение. Пусть случайная величина ξ – температура в квартире. Математическое ожидание этой случайной величины равно $M\xi = 20$, а дисперсия $D\xi = \sigma^2 = 2^2 = 4$.

1) После перехода к противоположному событию по неравенству Чебышева при $\varepsilon = 4$, получим:

$$P\left(\left|\xi - 20\right| \leq 5\right) \geq 1 - \frac{4}{5^2} = 0{,}84.$$

2) Предполагая, что случайная величина ξ распределена по нормальному закону, получим:

$$P\left(\left|\xi - 20\right| \leq 5\right) = \Phi\left(\frac{5}{2}\right) = \Phi\left(2{,}5\right) = 0{,}9876.$$

Из полученных результатов видно, что полученное точное значение вероятности события не противоречит ее оценки, полученной по неравенству Чебышева, ибо $0{,}9876 > 0{,}84$.

§ 6. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ ПО ТЕМЕ «ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ»

12. Суточная потребность электроэнергии в населенном пункте является случайной величиной с математическим ожиданием 2000 кВт/ч и дисперсией 20000. Используя неравенство Чебышева, оценить вероятность того, что в ближайший день расход электроэнергии в населенном пункте будет от 1500 до 2500 кВт/ч.

13. Среднее значение длины детали равно 50 см. Пользуясь леммой Чебышева, оценить вероятность того, что случайно взятая деталь окажется по длине:

а) более 49,5 см;

б) не более 50,5 см.

14. Дневная выручка магазина является случайной величиной со средним значением 10000 руб. и средним квадратическим отклонением 2000 руб.

1) С помощью неравенства Чебышева оценить вероятность того, что дневная выручка будет находиться в пределах от 6000 до 14000 руб.

2) Найти вероятность того же события, учитывая, что дневная выручка магазина является случайной величиной, распределенной по нормальному закону.

3) Объяснить различие результатов.

§ 7. ОЦЕНКА ПАРАМЕТРОВ

Пример 10. По схеме собственно-случайной бесповторной выборки в некотором регионе проведено обследование 150 строительных организаций из 2000 имеющихся с целью изучения объема производимых ими строительных работ. Объем строительных работ оценивается в млн. руб. Результаты обследования приведены в таблице 1:

259	257	245	255	235	256	240	226	252	254
246	247	262	248	228	245	255	242	250	269
243	241	245	256	258	217	263	240	241	261
246	248	243	242	263	257	261	255	252	255
239	254	247	252	246	255	242	240	261	253
255	256	239	242	251	260	249	240	243	257
243	255	258	255	253	242	251	244	272	248
251	231	270	241	245	246	240	254	231	228
249	243	253	242	242	249	255	252	237	241
239	239	256	238	257	244	244	235	256	245
254	246	237	238	264	237	261	259	242	264
262	258	234	257	250	266	243	266	263	262
239	260	253	248	249	261	246	245	267	258
249	259	239	241	248	254	241	234	247	238
244	246	247	264	250	252	243	248	261	227

Табл. 1

Выбрав соответствующее число промежутков, записать интервальный вариационный ряд, построить соответствующую ему гистограмму, полигон частот и кумулятивную кривую.

Решение. Для построения вариационного ряда необходимо произвести ранжирование исходных данных – упорядочивание их в порядке возрастания. Такая операция позволяет нам определить наименьшее $x_{\min}$ и наибольшее $x_{\max}$ значения среди представленных данных, а также произвести их группировку. В нашем случае наименьшее и наибольшее значения оказались равными $x_{\min} = 217$ и $x_{\max} = 272$.

Для разбиения представленных данных на отдельные группы или интервалы найдем число интервалов, которое определим по формуле Стерджеса

$$m = 1 + [3{,}322 \cdot lg\, n] = 1 + [3{,}322 \cdot lg\, 150] = 8,$$

где квадратные скобки $[x]$ обозначают целую часть числа.

Длину интервалов разбиения возьмем одинаковую и определим по формуле:

$$h = \frac{x_{\max} - x_{\min}}{m} = \frac{272 - 217}{8} = 6{,}875 \approx 7.$$

При расчете длины интервалов разбиения было произведено округление длины интервала до целого числа, ибо все исходные данные представляют собой целые числа. В общем случае при определении длины интервалов и для упрощения дальнейших вычислений целесообразно округлять длину интервала в большую сторону так, чтобы после запятой остался на один знак больше, чем у чисел в исходных данных.

Построим границы интервалов. Для этого определим величину Δ, равную $\Delta = x_{\min} + mh - x_{\max}$. Тогда границы искомых интервалов $\left(a_{i-1}, a_i\right)$ можно рассчитать по правилу: левая граница первого интервала будет равна $a_0 = x_{min} - \frac{\Delta}{2}$, а далее $a_i = a_{i-1} + h$ при $i = 1, 2, \ldots, m$. При таком выборе правая граница последнего интервала будет равна $a_m = x_{\max} + \frac{\Delta}{2}$.

В нашем случае $\Delta = 217 + 8 \cdot 7 - 272 = 1$ и границы интервалов будут следующими. Первый интервал будет иметь начало, равное

$a_0 = 217 - \frac{1}{2} = 216,5$, а его конец $a_1 = 216,5 + 7 = 223,5$. Начало второго интервала будет совпадать с концом первого, а его конец $a_2 = 223,5 + 7 = 230,5$. И так далее находим границы всех 8 интервалов. Очевидно, что $a_8 = 272,5$.

Далее, для каждого интервала определяется его срединное значение, как среднее арифметическое его концов. Так, например, для первого интервала срединное значение будет равно $\frac{216,5 + 223,5}{2} = 220$.

Определив границы интервалов, можно найти для каждого интервала число данных, принадлежащих ему.

Каждое из представленных в табл. 1 значений изучаемой случайной величины попадет только в один из интервалов разбиения. Подсчитав их количество, приходим к следующему вариационному ряду (табл. 2):

Интервалы, (a_{i-1}; a_i)	216,5 – 223,5	223,5 – 230,5	230,5 – 237,5	237,5 – 244,5	244,5 – 251,5	251,5 – 258,5	258,5 – 265,5	265,5 – 272,5	**Итого:**
Срединные значения, x_i	220	227	234	241	248	255	262	269	
Частоты, n_i	1	4	9	39	34	37	20	6	**150**
Накопленные частоты, $n_{x,i}$	1	5	14	53	87	124	144	150	

Табл. 2

В данной таблице кроме частот отображены накопленные частоты, определяемые как сумма частот вариант, не превышающих данного варианта.

Для первого интервала накопленная частота совпадает просто с частотой.

Для каждого последующего интервала накопленная частота равна сумме текущей частоты и накопленной частоты предыдущего интервала. Так, например, для второго интервала $\left(a_1, a_2\right)$ накопленная частота будет равна 4+1=5, для третьего интервала 9+5=14 и так далее.

Для графического изображения полученного интервального вариационного ряда построим *гистограмму*.

На каждом интервале построим столбик, высота которого равна частоте попадания в этот интервал (рис. 2).

Чтобы получить *полигон* того же распределения, соединим середины верхних оснований прямоугольников отрезками прямых (рис. 2).

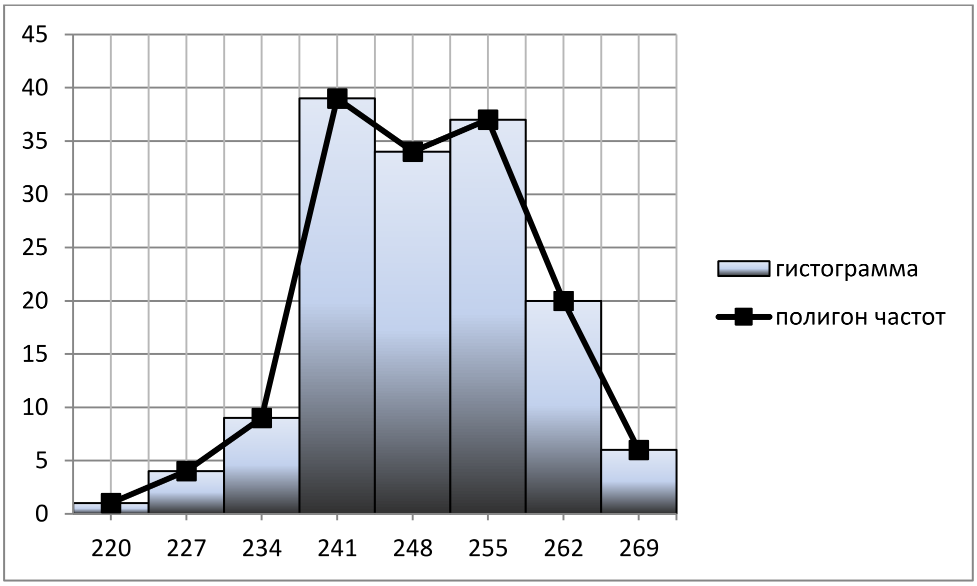

Рис. 2

Построение кумулятивной кривой начинают с точки, абсцисса которой равна началу первого интервала, а ордината равна нулю. Абсциссы следующих точек этой ломаной будут совпадать с концами интервалов, а ординаты – соответствующим им накопленным частотам (рис. 3).

Пример 11. Для вариационного ряда, построенного в **примере 10** (табл. 2), определить следующие выборочные характеристики:

а) моду,

б) медиану,

в) среднюю арифметическую, дисперсию и среднее квадратическое отклонение, г) долю строительных организаций, объем производимых работ которых не превышает величины, равной правой границе третьего интервала разбиения.

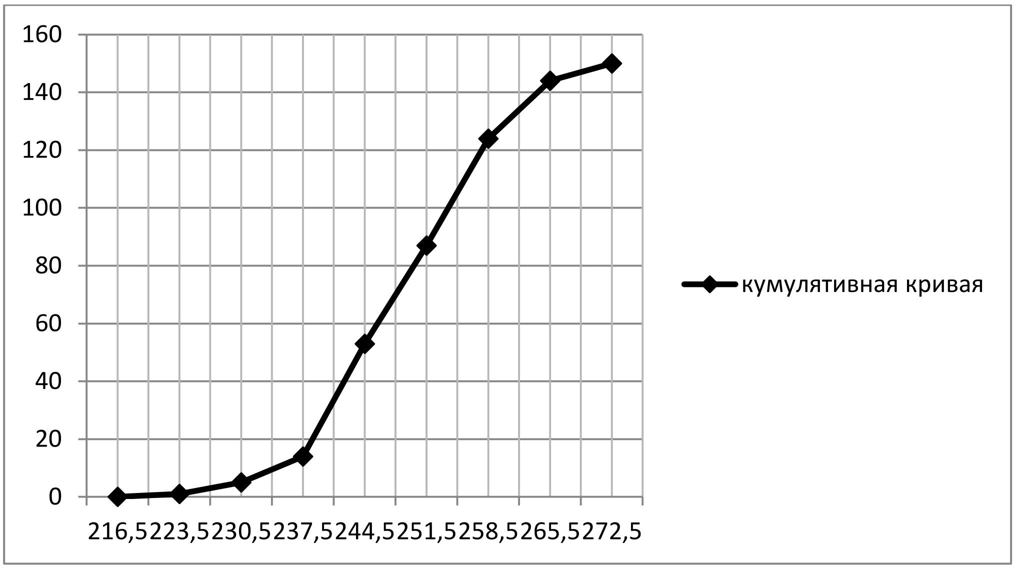

Рис. 3

Решение. а) Если интервальный вариационный ряд имеет только один интервал с наибольшей частотой, то этот интервал будем называть модальным. Если два соседних интервала имеют одинаковую наибольшую частоту, то их объединяют в один модальный интервал. В качестве приближенного значения моды берут его середину. Точное значение моды можно получить по формуле:

$$Mo = x_{м} + h \cdot \frac{n_{м} - n_{м-1}}{(n_{м} - n_{м-1}) + (n_{м} - n_{м+1})},$$

где $x_{м}$ и $n_{м}$ – соответственно начальное значение и частота модального интервала; $n_{м-1}$ и $n_{м+1}$ – соответственно частота интервала, предшествующего и следующего за модальным интервалом; h – длина модального интервала.

В нашем случае наибольшую частоту имеет четвертый интервал $(237{,}5;\ 244{,}5)$, и она равна 39. Поэтому приближенное значение моды будет равно середине интервала, то есть $Mo \approx 241$, а точное значение

$$Mo = 237{,}5 + 7 \cdot \frac{39-9}{(39-9)+(39-34)} = 243{,}5.$$

б) Для определения медианы интервального вариационного ряда находится медианный интервал, на который приходится середина ряда, то есть первый интервал, где сумма накопленных частот превышает половину наблюдений от общего объема выборки. Тогда

$$Me = x_e + \frac{1}{n_e}\left(\frac{n}{2} - n_{x,e-1}\right) \cdot h,$$

где x_e, n_e, h – соответственно начало, частота и ширина медианного интервала; $n_{x,e-1}$ – накопленная на предыдущих интервалах частота; n – объем выборки.

Для представленного вариационного ряда получим:

$$Me = 244{,}5 + \frac{1}{34}\left(\frac{150}{2} - 53\right) \cdot 7 = 249{,}03.$$

в) Для интервального вариационного ряда средняя арифметическая и средняя арифметическая квадратов вариант вариационного ряда вычисляются по формулам:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} x_i n_i \ ,$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{m} x_i^2 n_i \ ,$$

где x_i – варианты вариационного ряда, равные срединным значениям интервалов разбиения; n_i – соответствующие им частоты; m – число интервалов разбиения,

Получим:

$$\bar{x} = \frac{1}{150}(220 \cdot 1 + 227 \cdot 4 + 234 \cdot 9 + 241 \cdot 39 + 248 \cdot 34 + 255 \cdot 37 + 262 \cdot 20 + 269 \cdot 6) = 249{,}03.$$

$$\overline{x^2} = \frac{1}{150}(220^2 \cdot 1 + 227^2 \cdot 4 + 234^2 \cdot 9 + 241^2 \cdot 39 + 248^2 \cdot 34 + 255^2 \cdot 37 + 262^2 \cdot 20 + 269^2 \cdot 6) = 62110{,}57.$$

Выборочная дисперсия будет равна:

$$s^2 = \overline{x^2} - \bar{x}^2 = 62110{,}57 - 249{,}03^2 = 94{,}63 \ ,$$

а среднее квадратическое отклонение:

$$s = \sqrt{s^2} = \sqrt{94{,}63} = 9{,}73 \ .$$

г) Долю строительных организаций, объем производимых работ которых не превышает величины, равной правой границе третьего интервала разбиения, определим, исходя из вариационного ряда. В нашем случае число объектов, для которых объем производимых строительных работ меньше правой границы третьего интервала, равной 237,5, будет равно накопленной частоте указанного интервала. Из табл. 2 следует, что накопленная частота равна 14. Таким образом, выборочная доля будет равна

$$w = \frac{n_{x_3}}{n} = \frac{14}{150} \approx 0{,}093 \ .$$

Пример 12. На основе статистических данных, приведенных в **примере 10**, найти вероятность того, что средний объем производимых строительных работ всех строительных организаций региона отличается от полученного по выборке, не более чем на 1,5 млн. руб. по абсолютной величине.

Решение. Вероятность того, что средний объем производимых строительных работ всех строительных организаций региона отличается от полученного по выборке, не более чем на 1,5 млн. руб. по абсолютной величине, представляет собой доверительную вероятность или надежность. Она определяется через среднюю квадратическую ошибку выборки по формуле:

$$\gamma = P\{|\bar{x} - \bar{x}_0| \le \Delta_x\} = \ \Phi\left(\frac{\Delta_x}{\sigma_x}\right),$$

где Δ_x – предельная ошибка выборки, σ_x – средняя квадратическая ошибка выборки, а $\Phi(t) -$ удвоенная функция Лапласа (приложение 2-2).

При оценке генеральной средней средняя квадратическая ошибка собственно-случайной *бесповторной* выборки достаточно большого объема определяется по формуле:

$$\sigma_x \approx \sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)}\ .$$

По условию **примера 10** объемы генеральной совокупности и выборки соответственно равны $N = 2000$ и $n = 150$. Подставляя в последнее соотношение эти числовые значения и вычисленное в **примере 11** выборочное значение дисперсии, получим:

$$\sigma_x \approx \sqrt{\frac{94{,}63}{150}\left(1-\frac{150}{2000}\right)} = 0{,}764\ .$$

По условию предельная ошибка выборки составляет Δ_x=1,5 млн.руб., следовательно, искомая доверительная вероятность будет равна

$$\gamma = \Phi\left(\frac{1{,}5}{0{,}764}\right) = \Phi(1{,}96) = 0{,}95\ .$$

Пример 13. На основе статистических данных, приведенных в **примере 10**, найти границы, в которых с вероятностью 0,9876 заключен средний объем выполняемых работ всех организаций.

Решение. Предельная ошибка *бесповторной* выборки определяет точность полученных результатов и находится как $\Delta = u \cdot \sigma'_{\bar{x}}$, где u – аргумент удвоенной функции Лапласа, соответствующий доверительной вероятности γ, т.е. $\gamma = \Phi(u)$. Интервальная оценка генеральной средней (доверительный интервал) имеет вид:

$$\bar{x} - \Delta \le \bar{x}_0 \le \bar{x} + \Delta\ ,$$

где $\bar{x}$ – выборочная средняя арифметическая.

Для заданной доверительной вероятности $\gamma = 0{,}9876$ по таблице удвоенных значений функции Лапласа (приложение 2-2) находим, что ее аргумент $u = 2{,}5$. Следовательно, $\Delta = 2{,}5 \cdot 0{,}764 = 1{,}91$ и искомый доверительный интервал для генеральной средней имеет вид:

$$249{,}03 - 1{,}91 \le \bar{x}_0 \le 249{,}03 + 1{,}91$$

или

$$247{,}12 \le \bar{x}_0 \le 250{,}94.$$

Пример 14. На основе статистических данных, приведенных в **примере 10**, определить необходимый объем бесповторной выборки, чтобы с вероятностью 0,9876 гарантировать, что средний объем строительных работ, полученный по выборке, отличается от генеральной средней не более, чем на 1,5 млн. руб.

Решение. Объем *бесповторной* выборки определим через соответствующий объем *повторной* выборки. Для определения объема повторной выборки, который необходим для того, чтобы гарантировать определенную точность оценки генеральной средней, задаваемую предельной ошибкой выборки Δ_x, при заданной надежности γ, используем формулу:

$$n \approx \frac{u^2 \cdot s^2}{\Delta^2} .$$

По условию задачи доверительная вероятность равна $\gamma = 0,9876$, что соответствует аргументу удвоенной функции Лапласа равному $u = 2,5$. Предельная ошибка выборки равна $\Delta_x = 1,5$. На основании этого объем повторной выборки приблизительно будет равен (округление производим до целого числа всегда в большую сторону):

$$n \approx \frac{2,5^2 \cdot 94,63}{1,5^2} \approx 263 .$$

Зная объем повторной выборки и объем генеральной совокупности, вычисляем объем бесповторной выборки по формуле:

$$n' = \frac{nN}{n+N} = \frac{263 \cdot 2000}{263 + 2000} \approx 233 .$$

Пример 15. На основе статистических данных, приведенных в **примере 10**, найти вероятность того, что доля строительных организаций, объем строительных работ которых не превышает 237,5 млн. руб., отличается от полученной по выборке доли не более чем на 5% по абсолютной величине.

Решение. На основании вариационного ряда, ранее было определено, что выборочная доля $w = 0,093$. Вероятность того, что доля строительных организаций, объем строительных работ которых не превышает 237,5 млн.руб., отличается от полученной по выборке доли не более, чем на 5% по абсолютной величине, определяется через среднюю квадратическую ошибку выборки.

Средняя квадратическая ошибка собственно-случайной *бесповторной* выборки при оценке генеральной доли, находится по формуле:

$$\sigma_w' \approx \sqrt{\frac{w(1-w)}{n} \cdot \left(1 - \frac{n}{N}\right)} ,$$

где w – выборочная доля. Следовательно, средняя квадратическая ошибка бесповторной выборки будет равна:

$$\sigma_w' \approx \sqrt{\frac{0{,}093 \cdot (1 - 0{,}093)}{150} \cdot \left(1 - \frac{150}{2000}\right)} = 0{,}023 .$$

Так же, как и при оценке генеральной средней, доверительную вероятность выборочной доли определим по формуле:

$$\gamma = P\{|w - p| \le \Delta\} = \Phi\left(\frac{\Delta}{\sigma_w'}\right) .$$

Предельную ошибку выборки, равную 5%, представляем в виде доли: $\Delta = 0{,}05,$ следовательно, искомая доверительная вероятность будет равна:

$$\gamma = \Phi\left(\frac{0{,}05}{0{,}023}\right) = \Phi(2{,}17) = 0{,}97 .$$

Пример 16. На основе статистических данных, приведенных в **примере 10**, найти границы, в которых с вероятностью 0,9876 заключена доля строительных организаций, объем строительных работ которых не превышает 237,5 млн. руб.

Решение. Учитывая, что $\gamma = \Phi(u) = 0{,}9876$, по таблице удвоенных значений функции Лапласа (приложение 2-2) найдем $u = 2{,}5$ и определим предельную ошибку бесповторной выборки для доли:

$$\Delta = u \cdot \sigma_w' = 2{,}5 \cdot 0{,}023 \approx 0{,}058 .$$

Теперь искомый доверительный интервал для генеральной доли определяется соотношением:

$$0{,}093 - 0{,}057 \leq p \leq 0{,}093 + 0{,}057$$

или

$$0{,}036 \leq p \leq 0{,}150 \ .$$

Пример 17. На основе статистических данных, приведенных в **примере 10**, определить такой объем бесповторной выборки, чтобы с вероятностью 0,9876 доля строительных организаций, объем строительных работ которых не превышает 237,5 млн. руб., отличалась от полученной по выборке не более чем на 5% (по абсолютной величине).

Ответить на тот же вопрос, если о выборочной доле ничего неизвестно.

Решение. Также как и ранее при определении объема выборки при оценке генеральной средней, найдем сначала объем повторной выборки. Объем повторной выборки при оценке генеральной доли определяется соотношением:

$$n \approx \frac{u^2 w(1-w)}{\Delta^2} \ .$$

В качестве выборочной доли возьмем состоятельную оценку $w = 0{,}093$, полученную ранее. Учитывая что $\Delta = 0{,}05$, $\gamma = \Phi(u) = 0{,}9876$ и $u = 2{,}5$, объем повторной выборки приблизительно будет равен:

$$n \approx \frac{2{,}5^2 \cdot 0{,}093 \cdot (1 - 0{,}093)}{0{,}05^2} \approx 211 \ .$$

Имея объемы повторной выборки и генеральной совокупности, определяем объем бесповторной выборки по формуле:

$$n' = \frac{nN}{n+N} = \frac{221 \cdot 2000}{221 + 2000} \approx 190 \ .$$

Рассмотрим случай, когда никаких предварительных исследований не проводилось и о выборочной доле ничего неизвестно. В этом случае можно вычислить максимально возможный объем повторной выборки, соответствующий заданной доверительной вероятности и точности, по формуле:

$$n = \frac{u^2}{4 \cdot \Delta^2} \ .$$

Подставляя числовые значения, получим:

$$n = \frac{2,5^2}{4 \cdot 0{,}05^2} = 625 \ ,$$

и, соответственно, объем бесповторной выборки будет равен:

$$n^{'} = \frac{nN}{n+N} = \frac{625 \cdot 2000}{625 + 2000} \approx 476 \ .$$

Очевидно, что максимально возможное значение объема выборки оказалось значительно больше необходимого. Это объясняется тем, что выборочная доля существенно отличается от значения равного 0,5, используемого при оценке максимального объема.

§ 8. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ ПО ТЕМЕ «ОЦЕНКА ПАРАМЕТРОВ»

15. Изучается успеваемость студентов некоторого вуза по математике. По схеме собственно-случайной бесповторной выборки из 500 студентов вуза было отобрано 50. Результаты опроса студентов представляют собой следующий набор чисел:

$$\begin{array}{c} 3;4;5;4;2;3;3;3;5;4;3;5;5;2;3;5;3;5;3;5;4;4;3;3;4; \\ \underbrace{3;3;3;4;3;4;3;5;3;4;4;3;5;3;3;5;4;2;5;3;4;2;3;5;4}_{n=50}. \end{array}$$

а) Построить дискретный вариационный ряд.

б) Построить полигон частот и кумулятивную кривую для данного ряда.

в) вычислить медиану, моду, среднюю арифметическую и дисперсию для данного вариационного ряда.

16. В мастерской по ремонту и обслуживанию бытовой радиоэлектронной аппаратуры было проведено обследование на предмет определения среднего числа обращений клиентов в день. По схеме бесповторной собственно-случайной выборки отобрано 50 рабочих дней прошедшего года (будем считать, что в году 300 рабочих дней). Получены следующие данные о числе вызовов в день:

Число вызовов в день	Менее 10	10 – 15	15 – 20	20 – 25	Более 25	Всего
Количество дней	6	13	18	10	3	50

Представить приведенные опытные данные графически: построить гистограмму, полигон частот и кумулятивную кривую.

Определить:

1) Выборочные оценки числовых характеристик случайной величины – числа вызовов в день: моду, медиану, среднюю арифметическую, дисперсию, а также долю дней в году, число вызовов в которые не превосходит 15.

2) Вероятность того, что среднее число вызовов в день в году отличается от среднего числа вызовов, полученного по выборке, не более чем на 1,5 по абсолютной величине.

3) Границы, в которых с вероятностью 0,9876 заключено среднее число вызовов в день в году.

4) Объем бесповторной выборки, чтобы с вероятностью 0,9876 среднее число вызовов в день полученное по выборке, отличалось от генеральной средней не более чем на 1.

5) Вероятность того, что доля дней в году, число вызовов в которых не превышает 15, отличается от полученной по выборке доли не более чем на 5% по абсолютной величине.

6) Границы, в которых с вероятностью 0,9545 заключена доля дней в году, в которых число вызовов в день не превышает 15.

7) Объем бесповторной выборки, чтобы с вероятностью 0,9545 доля дней в году, число вызовов в которые не более 20, отличалась от полученной по выборке не более чем на 5% (по абсолютной величине).

Ответить на последний вопрос, если о выборочной доле ничего неизвестно.

§ 9. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Пример 18. Для выборки, приведенной в **примере 10**, подобрать соответствующее теоретическое распределение и на уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что изучаемая случайная величина – объем строительных работ, производимых строительными организациями некоторого региона – распределена по этому закону. Если гипотеза о виде закона распределения не будет отвергнута, построить на том же рисунке, что и гистограмма, выровненную теоретическую функцию плотности распределения.

Решение. По виду гистограммы распределения объема производимых строительных работ (рис. 2) можно предположить, что изучаемая случайная величина распределена по нормальному закону. Проверку этой гипотезы можно провести, используя критерий согласия χ^2 –Пирсона.

Наблюдаемое значение статистики определяется по эмпирическим и теоретическим частотам по формуле:

$$\chi^2_{\text{набл.}} = \sum_{i=1}^{m} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i} \,,$$

где n_i – эмпирические, а $n \cdot p_i$ – теоретические частоты.

Для определения теоретических частот нам нужны параметры этого закона распределения, а именно – математическое ожидание a и среднее квадратическое отклонение σ. Точные значения этих параметров теоретического закона распределения нам неизвестны, поэтому заменим их оценками, полученными по выборке.

В **примере 11** были рассчитаны выборочная средняя, выборочная дисперсия и среднее квадратическое отклонение. Поэтому будем считать, что

$$a \approx \bar{x} = 249{,}03 \ , \qquad \sigma \approx s = 9{,}73 \ .$$

Теоретические вероятности попадания в рассматриваемые интервалы для нормально распределенной случайной величины выражаются через удвоенную функцию Лапласа по формуле:

$$p_i = P(x_i \le \xi < x_{i+1}) = \frac{1}{2}\left(\Phi\left(\frac{x_{i+1} - \bar{x}}{s}\right) - \Phi\left(\frac{x_i - \bar{x}}{s}\right)\right) \ , \quad i = 1,\ 2,\ \ldots,\ m.$$

Так как для нормально распределенной случайной величины возможные значения могут теоретически принимать любые значения на всей числовой оси, то для корректности вычислений целесообразно левую границу первого интервала принять равной $-\infty$, а правую границу последнего интервала следует взять равной $+\infty$.

Так, например, вероятность попадания в первый и второй интервалы будут соответственно равны:

$$p_1 = P(-\infty \le \xi < 223{,}5) =$$

$$= \frac{1}{2}\left(\Phi\left(\frac{223{,}5 - 249{,}03}{9{,}73}\right) - \Phi\left(\frac{-\infty - 249{,}03}{9{,}73}\right)\right) =$$

$$= \frac{1}{2}\big(\Phi(-2{,}62) - \Phi(-\infty)\big) =$$

$$= \frac{1}{2}(-0{,}9912 + 1) = 0{,}0044;$$

$$p_2 = P(223{,}5 \le \xi < 230{,}5) =$$

$$= \frac{1}{2}\left(\Phi\left(\frac{230{,}5 - 249{,}03}{9{,}73}\right) - \Phi\left(\frac{223{,}5 - 249{,}03}{9{,}73}\right)\right) =$$

$$= \frac{1}{2}\big(\Phi(-1{,}90) - \Phi(-2{,}62)\big) =$$

$$= \frac{1}{2}(-0{,}9426 + 0{,}9912) = 0{,}0243.$$

В последних вычислениях значения удвоенной функции Лапласа определяются по таблице (приложение 2-2). Дальнейшие вычисления удобно представить в виде таблицы:

x_i	x_{i+1}	n_i	$\frac{x_i-\bar{x}}{s}$	$\frac{x_{i+1}-\bar{x}}{s}$	$\Phi\left(\frac{x_i-\bar{x}}{s}\right)$	$\Phi\left(\frac{x_{i+1}-\bar{x}}{s}\right)$	p_i	$n\cdot p_i$	$\frac{(n_i-n\cdot p_i)^2}{n\cdot p_i}$
$-\infty$	223,5	$\left.\begin{matrix}1\\4\end{matrix}\right\}5$	$-\infty$	–2,62	–1	–0,9912	0,00440	$\left.\begin{matrix}0{,}66\\3{,}65\end{matrix}\right\}4{,}31$	0,112
223,5	230,5		–2,62	–1,90	–0,9912	–0,9426	0,02430		
230,5	237,5	9	–1,90	–1,18	–0,9426	–0,7620	0,09030	13,55	1,525
237,5	244,5	39	–1,18	–0,47	–0,7620	–0,3612	0,20040	30,06	2,659
244,5	251,5	34	–0,47	0,25	–0,3612	0,1974	0,27930	41,90	1,488
251,5	258,5	37	0,25	0,97	0,1974	0,6679	0,23525	35,29	0,083
258,5	265,5	20	0,97	1,69	0,6679	0,9090	0,12055	18,08	0,203
265,5	$+\infty$	6	1,69	$+\infty$	0,9090	1	0,04550	6,83	0,100
$\sum$		150					1	150	**6,170**

При расчете статистики χ^2 необходимо иметь в виду, что для корректного применения критерия Пирсона в каждом интервале должно быть количество наблюдений не менее пяти. Если в каком-либо интервале число наблюдений меньше пяти, то его имеет смысл объединить с соседним интервалом, так чтобы в объединенных интервалах число наблюдений было не менее пяти.

Учитывая вышесказанное, объединим первый и второй интервалы, ибо количество наблюдений, попавших в них меньше пяти. В результате получим, что наблюдаемое значение статистики $\chi^2_{набл.} = 6{,}170$.

Для нахождения критического значения статистики, вычислим число степеней свободы. При этом учтем, что два первых интервала объединялись и, таким образом, число интервалов m = 7, и что для нормального закона распределения число параметров распределения r =2. Таким образом, в нашем случае число степеней свободы будет равно $k = 7 - 2 - 1 = 4$ и по таблице значений χ^2 – критерия Пирсона (приложение 6) находим, что $\chi^2_{кр.}(0,05;4)=9,49$.

Сравниваем вычисленное по выборке значение статистики $\chi^2_{набл.}$ с критическим. Так как вычисленная по выборке статистика не превосходит критического значения (6,170 < 9,49), то это позволяет утверждать, что опытные данные на заданном уровне значимости не противоречат гипотезе о выбранном нормальном законе распределения изучаемой случайной величины, или, другими словами, выборочные данные согласуются с выдвинутой нулевой гипотезой.

Для графического изображения теоретической плотности распределения на графике, изображающем гистограмму эмпирического распределения, необходимо использовать одинаковый масштаб по оси ординат.

Так как при построении гистограммы эмпирического распределения по оси ординат откладывались эмпирические частоты, теоретическую кривую функции плотности распределения $\varphi(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$ необходимо растянуть (сжать) вдоль оси ординат в nh раз, где n – объем выборки, h – длина интервалов разбиения, то есть построить кривую $\varphi_1(x)=\frac{nh}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$. Выравнивающую теоретическую кривую $\varphi_1(x)$ можно приближенно построить по точкам (x_i, np_i), где в качестве x_i целесообразно взять середины интервалов разбиения, а np_i – соответствующие им теоретические частоты. Максимальное значение такой функции $\varphi_1(x)$ достигается в точке $x=a\approx\bar{x}$ и равно:

$$\max\varphi_1(x)=\varphi_1(\bar{x})=\frac{nh}{\sigma\sqrt{2\pi}}\approx\frac{0,3989nh}{s}.$$

Для рассматриваемого примера $\max \varphi_1(x) \approx \dfrac{0{,}3989 \cdot 150 \cdot 7}{9{,}73} = 43{,}05$. График гистограммы и теоретической кривой представлен на рис. 4.

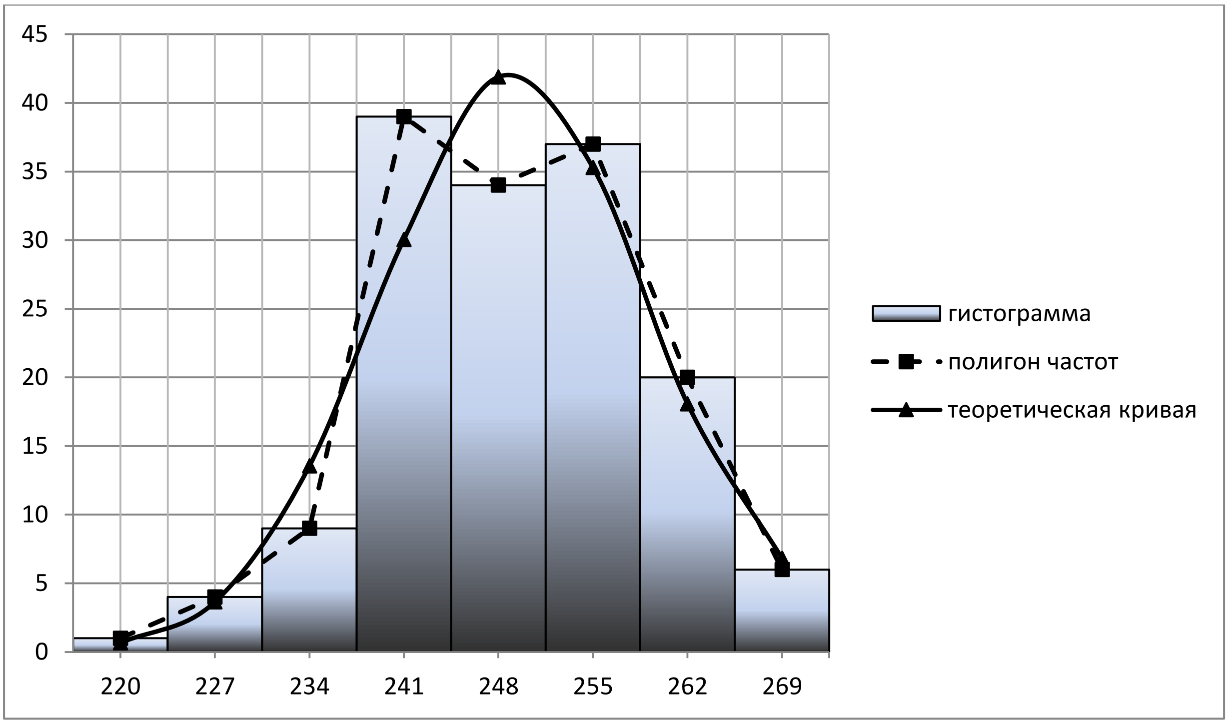

Рис. 4

Пример 18. Имеются статистические данные о времени ожидания трамвая в некотором населенном пункте, представленные в таблице:

Время ожидания, мин.	0 – 1	1 – 2	2 – 3	3 – 4	4 – 5	5 – 6	6 – 7	7 – 8	Итого
Срединные значения	0,5	1,5	2,5	3,5	4,5	5,5	6,5	7,5	
Частота	11	15	12	15	11	15	14	7	**100**

На основании вида гистограммы и полигона частот, подобрать соответствующее теоретическое распределение и на уровне значимости 0,05 проверить гипотезу о том, что время ожидания трамвая распределено по выбранному закону. Если гипотеза о виде закона распределения не будет отвергнута, построить на графике гистограммы выровненную теоретическую функцию плотности распределения.

Решение. Начнем решение этой задачи с построения гистограммы и полигона частот рис. 5.

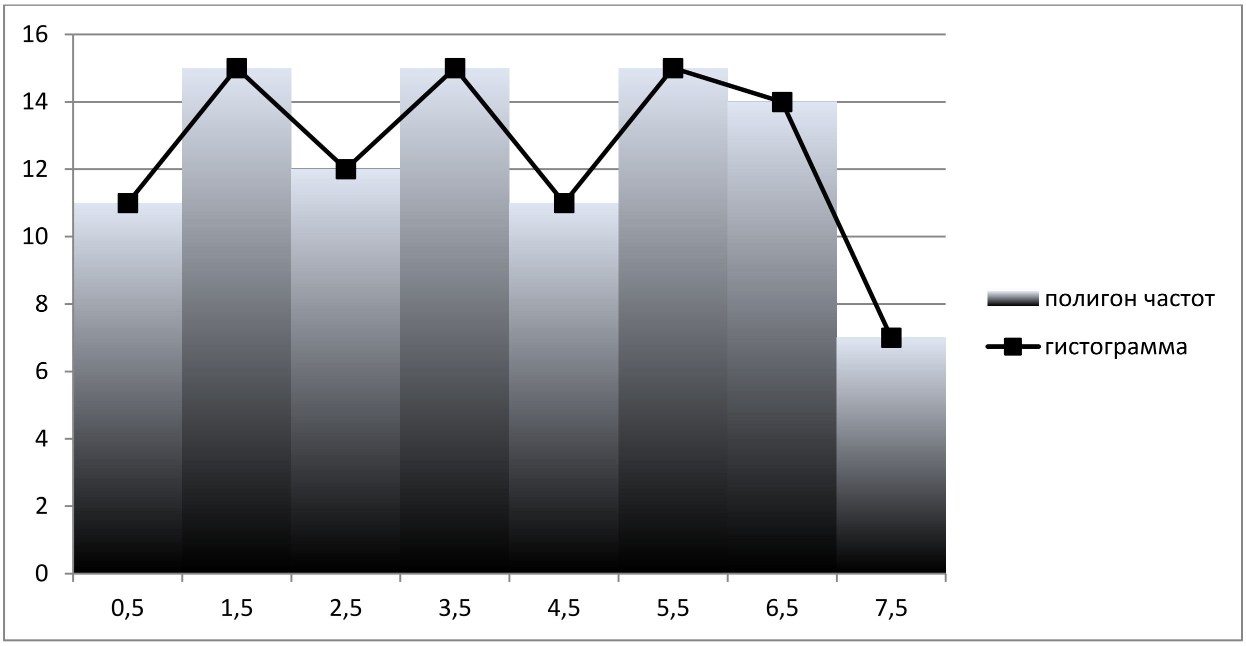

Рис. 5

Исходя из вида гистограммы можно предположить, что рассматриваемая случайная величина имеет равномерное распределение.

Так как все опытные данные принадлежат отрезку [0; 8], то можем считать, что и функция плотности распределения также отлична от нуля только на этом промежутке, следовательно, функция плотности примет вид:

$$\varphi(x)=\begin{cases}\dfrac{1}{8-0}=0{,}125 & \textit{при}\quad 0\le x\le 8,\\ 0 & \textit{в остальных случаях.}\end{cases}$$

Все интервалы $\left[x_i, x_{i+1}\right)$ имеют одну длину, равную 1, следовательно, все теоретические вероятности попадания в эти интервалы будут одинаковыми и равными $p_i = 1\cdot 0{,}125 = 0{,}125$.

Также будут одинаковы и соответствующие им теоретические частоты $np_i = 100 \cdot 0{,}125 = 12{,}5$. Дальнейшие расчеты представлены в таблице:

x_i	x_{i+1}	n_i	p_i	$n \cdot p_i$	$\dfrac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
0	1	11	0,125	12,5	0,18
1	2	15	0,125	12,5	0,5
2	3	12	0,125	12,5	0,02
3	4	15	0,125	12,5	0,5
4	5	11	0,125	12,5	0,18
5	6	15	0,125	12,5	0,5
6	7	14	0,125	12,5	0,18
7	8	7	0,125	12,5	2,42
$\sum$		100	1	100	$4{,}48 = \chi^2_{набл.}$

При уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $k = 8 - 2 - 1 = 5$ (равномерный закон распределения имеет два параметра распределения) из приложения 6 критическое значение статистики χ^2 – Пирсона:

$$\chi^2_{кр.}(0{,}05;6) = 11{,}1.$$

Таким образом, значение статистики $\chi^2_{набл.}$, вычисленное по выборке, не превосходит критического значения: 4,48 < 11,1. Последнее позволяет утверждать, что при заданном уровне значимости опытные данные не противоречат гипотезе о равномерном законе распределения, или опытные данные согласуются с выдвинутой нулевой гипотезой.

Построим выравнивающую теоретическую плотность распределения. Вновь учтем, что при построении гистограммы эмпирического распределения по оси ординат откладывались эмпирические частоты, и вместо теоретической кривой функции плотности распределения $\varphi(x)$ будем строить кривую $\varphi_1(x) = nh \cdot \varphi(x)$. Выравнивающую теоретическую кривую $\varphi_1(x)$ можно приближенно построить по точкам (x_i, np_i), где в качестве x_i целесообразно взять середины интервалов разбиения, а np_i– соответствующие им теоретические частоты. Очевидно, что эта функция на отрезке от 0 до 8 будет представлять собой отрезок прямой, параллельный оси абсцисс (рис. 6). Вне этого отрезка функция будет равна нулю.

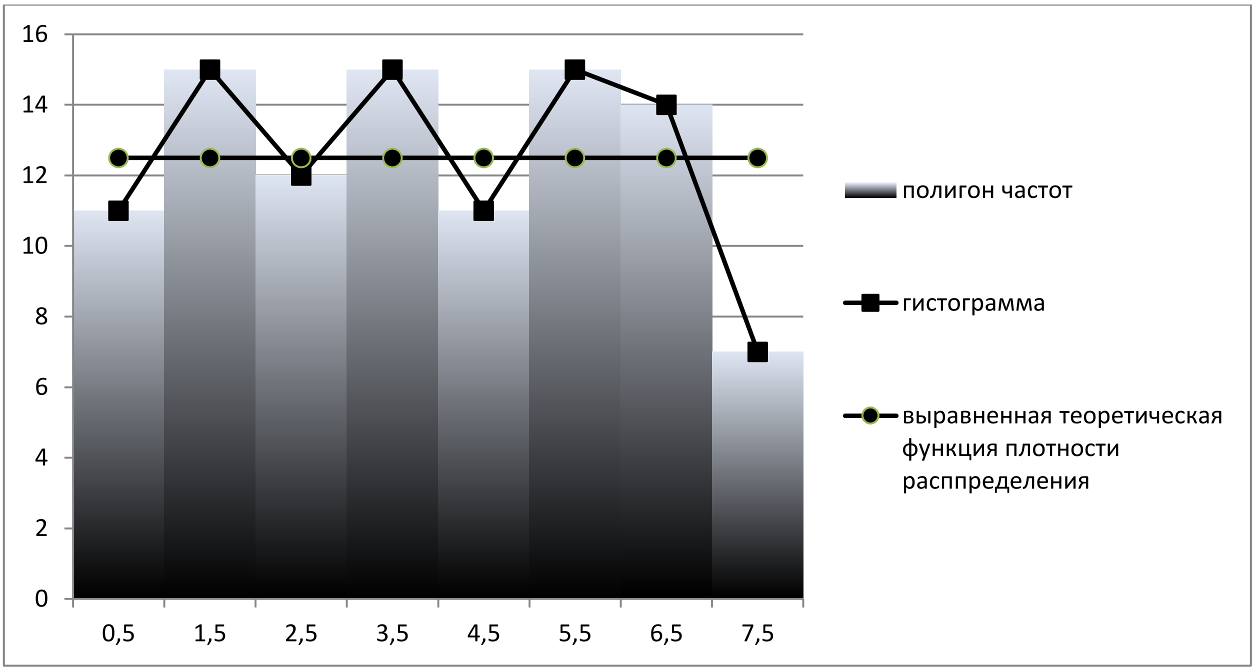

Рис. 6

§ 10. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ ПО ТЕМЕ «ПРОВЕРКА СТАТИЧТИЧЕСКИХ ГИПОТЕЗ»

17. Автомобиль подъезжает к перекрестку, регулируемому светофором, в некоторый момент времени. На светофоре – красный сигнал. Исследовалось время, в течение которого водителю автомобиля приходится ждать зеленого сигнала светофора. Проведено 100 независимых наблюдений. Эмпирические данные представлены в виде вариационного ряда:

Время ожидания, сек.	Менее 5	5 – 10	10 – 15	15 – 20	20 – 25	Более 25	Итого
Частота	19	15	18	15	13	20	**100**

На уровне значимости 0,05 проверить гипотезу о том, что время ожидания зеленого сигнала светофора представляет собой случайную величину, равномерно распределенную на отрезке [0, 30].

18. На предприятии работает 2000 сотрудников. Для изучения стажа работы сотрудников на этом предприятии по схеме собственно случайной бесповторной выборки отобрано 400 человек. Полученные данные о стаже работы представлены в таблице:

Стаж работы, годы	До 5	5–10	10–15	15–20	Свыше 20	Итого
Число работников	32	56	92	120	100	400

На уровне значимости 0,05 проверить гипотезу о том, что стаж работников предприятия распределен по нормальному закону.

19. Для вариационного ряда, приведенного в **задаче 18**, на уровне значимости 0,05 проверить гипотезу о том, что стаж работников предприятия распределен равномерно на отрезке [5, 20].

§ 11. ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ

Пример 20. Проведено обследование 70 российских коммерческих банков на предмет изучения зависимости их прибыли от объема вложений в ценные бумаги. Получено следующее распределение банков по объему вложений в ценные бумаги ξ (млн. руб.) и полученной от них прибыли η (млн. руб.):

η / ξ	10 – 12	12 – 14	14 – 16	16 – 18	18 – 20	20 – 22	Итого:
100 – 130	2	1	3				6
130 – 160	1	2	5	3	2		13
160 – 190		2	7	8	7	3	27
190 – 220			2	7	5	2	16
220 – 250				5	2	1	8
Итого:	3	5	17	23	16	6	**70**

1. Вычислить групповые средние $\overline{x}_i$ и $\overline{y}_j$, построить поле корреляции и эмпирические линии регрессии;

2. Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

б) используя соответствующее уравнение регрессии, оценить полученную прибыль при объеме вложений в ценные бумаги 150 млн. руб.;

в) вычислить коэффициент корреляции, сделать вывод о тесноте и направлении связи между переменными ξ и η;

г) на уровне значимости α = 0,05 с помощью критерия Стьюдента оценить значимость коэффициента корреляции.

Решение. 1. Полученные в ходе обследования банков эмпирические данные представляют собой двумерную выборку, объем которой равен 70. По каждой переменной они представляют собой интервальный вариационный ряд. Для упрощения дальнейшей обработки заменим интервальные вариационные ряды их дискретными аналогами. Для этого каждый интервал разбиения, как по переменной ξ, так и по переменной η, будем характеризовать их срединным значением.

Вычислим групповые средние. Для каждого срединного значения x_i вычислим $\bar{y}_i$ и для каждого значения y_j вычислим $\bar{x}_j$ по следующим формулам:

$$\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{m} y_j n_{ij}, \quad \bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{l} x_i n_{ij}.$$

Имеем:

$$\bar{y}_1 = \frac{1}{6}(11 \cdot 2 + 13 \cdot 1 + 15 \cdot 3) = 13{,}33;$$

$$\bar{y}_2 = \frac{1}{13}(11 \cdot 1 + 13 \cdot 2 + 15 \cdot 5 + 17 \cdot 3 + 19 \cdot 2) = 15{,}46;$$

$$\bar{y}_3 = \frac{1}{27}(13 \cdot 2 + 15 \cdot 7 + 17 \cdot 8 + 19 \cdot 7 + 21 \cdot 3) = 17{,}15;$$

$$\bar{y}_4 = \frac{1}{16}(15 \cdot 2 + 17 \cdot 7 + 19 \cdot 5 + 21 \cdot 2) = 17{,}88;$$

$$\bar{y}_5 = \frac{1}{8}(17 \cdot 5 + 19 \cdot 2 + 21 \cdot 1) = 18{,}00.$$

$$\bar{x}_1 = \tfrac{1}{3}(115 \cdot 2 + 145 \cdot 1) = 125{,}00;$$

$$\bar{x}_2 = \tfrac{1}{5}(115 \cdot 1 + 145 \cdot 2 + 175 \cdot 2) = 151{,}00;$$

$$\bar{x}_3 = \tfrac{1}{17}(115 \cdot 3 + 145 \cdot 5 + 175 \cdot 7 + 205 \cdot 2) = 159{,}12;$$

$$\bar{x}_4 = \tfrac{1}{23}(145 \cdot 3 + 175 \cdot 8 + 205 \cdot 7 + 235 \cdot 5) = 193{,}26;$$

$$\bar{x}_5 = \tfrac{1}{16}(145 \cdot 2 + 175 \cdot 7 + 205 \cdot 5 + 235 \cdot 2) = 188{,}13;$$

$$\bar{x}_6 = \tfrac{1}{6}(175 \cdot 3 + 205 \cdot 2 + 235 \cdot 1) = 195{,}00.$$

Запишем все полученные результаты в виде корреляционной таблицы, в которой вычисленные значения групповых средних $\bar{x}_j$ приведены в последней строке, а $\bar{y}_i$ – в последнем столбце:

ξ	*Середины интер-валов*	η						n_i	$\bar{y}_j$
		10 – 12	12 – 14	14 – 16	16 – 18	18 – 20	20 – 22		
	x_i \ y_j	11	13	15	17	19	21		
100 – 130	115	2	1	3				6	13,33
130 – 160	145	1	2	5	3	2		13	15,46
160 – 190	175		2	7	8	7	3	27	17,15
190 – 220	205			2	7	5	2	16	17,88
220 – 250	235				5	2	1	8	18,00
	n_j	3	5	17	23	16	6	**70**	
	$\bar{x}_i$	125,00	15,00	159,12	193,26	188,13	195,00		

На основании данных, представленных в корреляционной таблице, строим поле корреляции и эмпирические линии регрессии (рис. 7). Поле корреляции – точки на плоскости $\left(x_i, y_j\right)$, для которых эмпирические частоты n_{ij} отличны от нуля. Эмпирическая линия регрессии η по ξ (или Y на X) по определению есть ломаная линия, соединяющая точки $\left(x_i, \bar{y}_i\right)$, а, соответственно эмпирическая линия регрессии ξ по η (или X на Y) – ломаная, соединяющая точки $\left(\bar{x}_j, y_j\right)$.

2. а) Найдем выборочные линейные уравнения регрессии. Уравнения регрессии η по ξ и ξ по η имеют вид:

$$y_x - \bar{y} = b_{\eta\zeta}(x - \bar{x}) \text{ , } x_y - \bar{x} = b_{\xi\eta}(y - \bar{y}) \text{ ,}$$

где $b_{\xi\eta} = \dfrac{\mu}{s_\eta^2}$ и $b_{\eta\xi} = \dfrac{\mu}{s_\xi^2}$ – выборочные коэффициенты регрессии ξ по η и ξ по η, соответственно; $s_\xi^2 = \overline{x^2} - \overline{x}^2$ и $s_\eta^2 = \overline{y^2} - \overline{y}^2$ выборочные дисперсии переменных ξ и η, соответственно; $\mu = \overline{xy} - \overline{x}\,\overline{y}$ – выборочный корреляционный момент или выборочная ковариация.

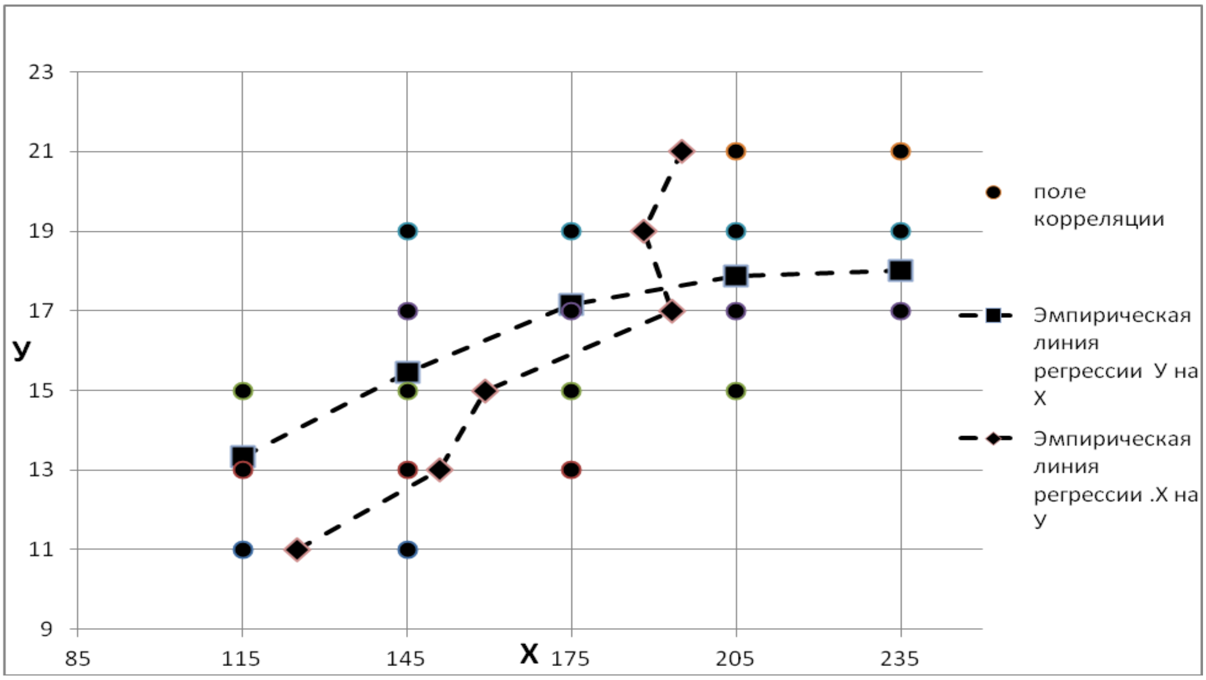

Рис. 7

Для вычисления всех коэффициентов найдем необходимые суммы:

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

$$= \frac{1}{70}(115 \cdot 6 + 145 \cdot 13 + 175 \cdot 27 + 205 \cdot 16 + 235 \cdot 8) = 178{,}0;$$

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{m} y_j n_j =$$

$$= \frac{1}{70}(11 \cdot 3 + 13 \cdot 5 + 15 \cdot 17 + 17 \cdot 23 + 19 \cdot 16 + 21 \cdot 6) = 16{,}77;$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{l} x_i^2 n_i =$$

$$= \frac{1}{70}(115^2 \cdot 6 + 145^2 \cdot 13 + 175^2 \cdot 27 + 205^2 \cdot 16 + 235^2 \cdot 8) =$$

32767,86;

$$\overline{y^2} = \frac{1}{n}\sum_{j=1}^{m} y_j^2 n_j =$$

$$= \frac{1}{70}(11^2 \cdot 3 + 13^2 \cdot 5 + 15^2 \cdot 17 + 17^2 \cdot 23 + 19^2 \cdot 16 + 21^2 \cdot 6) =$$

287,17;

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{l}\sum_{j=1}^{m} x_i y_j n_{ij} =$$

$$= \frac{1}{70}[115 \cdot (11 \cdot 2 + 13 \cdot 1 + 15 \cdot 3) +$$

$$+145 \cdot (11 \cdot 1 + 13 \cdot 2 + 15 \cdot 5 + 17 \cdot 3 + 19 \cdot 2) +$$

$$+175 \cdot (13 \cdot 2 + 15 \cdot 7 + 17 \cdot 8 + 19 \cdot 7 + 21 \cdot 3) +$$

$$+205 \cdot (15 \cdot 2 + 17 \cdot 7 + 19 \cdot 5 + 21 \cdot 2) +$$

$$+235 \cdot (17 \cdot 5 + 19 \cdot 2 + 21 \cdot 1)] = 3026{,}29.$$

Таким образом, имеем:

$$\mu = 3026,29 - 178 \cdot 17,77 = 40,97.$$

$$\bar{s}_{\xi}^{2} = 32767,86 - 178^2 = 1083,86; \quad \bar{s}_{\eta}^{2} = 287,17 - 16,77^2 = 5,89;$$

$$b_{\eta\xi} = \frac{40,97}{1083,86} = 0,038; \quad b_{\xi\eta} = \frac{40,97}{5,89} = 6,96.$$

Итак, уравнения регрессии имеют вид:

$$y_x - 16,77 = 0,038 \cdot (x - 178,0)$$

$$x_y - 178,0 = 6,96 \cdot (y - 16,77)$$

или

$$y_x = 0,038x + 10,0;$$

$$x_y = 6,96y + 61,28.$$

Построим графики полученных линий регрессии на том же самом чертеже, где построены эмпирические линии регрессии. Для построения прямой линии достаточно взять две точки, удовлетворяющие ее уравнению.

Заметим, что у нас уже есть одна точка, которая удовлетворяет обоим уравнениям регрессии, и координаты которой представляют собой средние значения признаков, то есть $(\overline{x}, \overline{y})$. Следовательно, это точка пересечения прямых линий регрессии. В нашем случае она имеет координаты $(178{,}0; 16{,}77)$.

Для построения каждой из линий регрессии выберем дополнительную точку.

В качестве второй точки для линии регрессии η по ξ, например, можно взять точку, абсцисса которой равна $x = 115$. Нетрудно вычислить соответствующую ей ординату: $y_x = 0{,}038 \cdot 115 + 10{,}0 = 14{,}37$. Проводим через точки $(115; 14{,}37)$ и $(178{,}0; 16{,}77)$ теоретическую прямую линию регрессии η по ξ (рис. 12). Аналогично, для построения теоретической линии регрессии ξ по η возьмем ординату $y = 11$. Для нее получим, что $x_y = 6{,}96 \cdot 11 + 61{,}28 = 137{,}84$. Следовательно, имеем еще одну точку $(11; 137{,}84)$. Проводим через нее и общую точку прямую линию. Графики эмпирических и теоретических линий представлены на рис.8.

б) Оценим полученную прибыль при объеме вложений в ценные бумаги 150 млн. руб. Для этого воспользуемся уравнением регрессии η по ξ, подставив в него вместо x значение равное 150. Получим:

$$y_x = 0{,}038 \cdot 150 + 10{,}0 = 15{,}7.$$

Экономический смысл полученного результата состоит в том, что при вложении в ценные бумаги 150 млн.руб., банки в среднем получают 15,7 млн.руб. прибыли.

в) Выборочный коэффициент линейной корреляции удобно вычислить по формуле:

$$r = \pm\sqrt{b_{\eta\xi} b_{\xi\eta}}.$$

В последнем соотношении знак «+» перед радикалом берется, если коэффициенты регрессии положительны, и знак «–», если коэффициенты регрессии отрицательные. В нашем случае коэффициенты регрессии положительны, и, следовательно, коэффициент корреляции также будет положителен и равен:

$$r = \sqrt{0{,}038 \cdot 6{,}96} = \sqrt{0{,}26448} \approx 0{,}51.$$

Поскольку полученный коэффициент корреляции положительный, то можно сделать вывод о том, что между рассматриваемыми переменными наблюдается прямая корреляционная связь. А так как он удовлетворяет соотношению $0{,}4 < r < 0{,}7$, связь можно считать умеренной.

г) Проверим значимость коэффициента корреляции в рассматриваемом примере. Последнее предполагает проверку нулевой гипотезы H_0 об отсутствии линейной корреляционной связи между переменными в генеральной совокупности, т.е. предполагается, что коэффициент корреляции в генеральной совокупности равен нулю

$$H_0 \colon \rho = 0,$$

где ρ – коэффициент корреляции в генеральной совокупности.

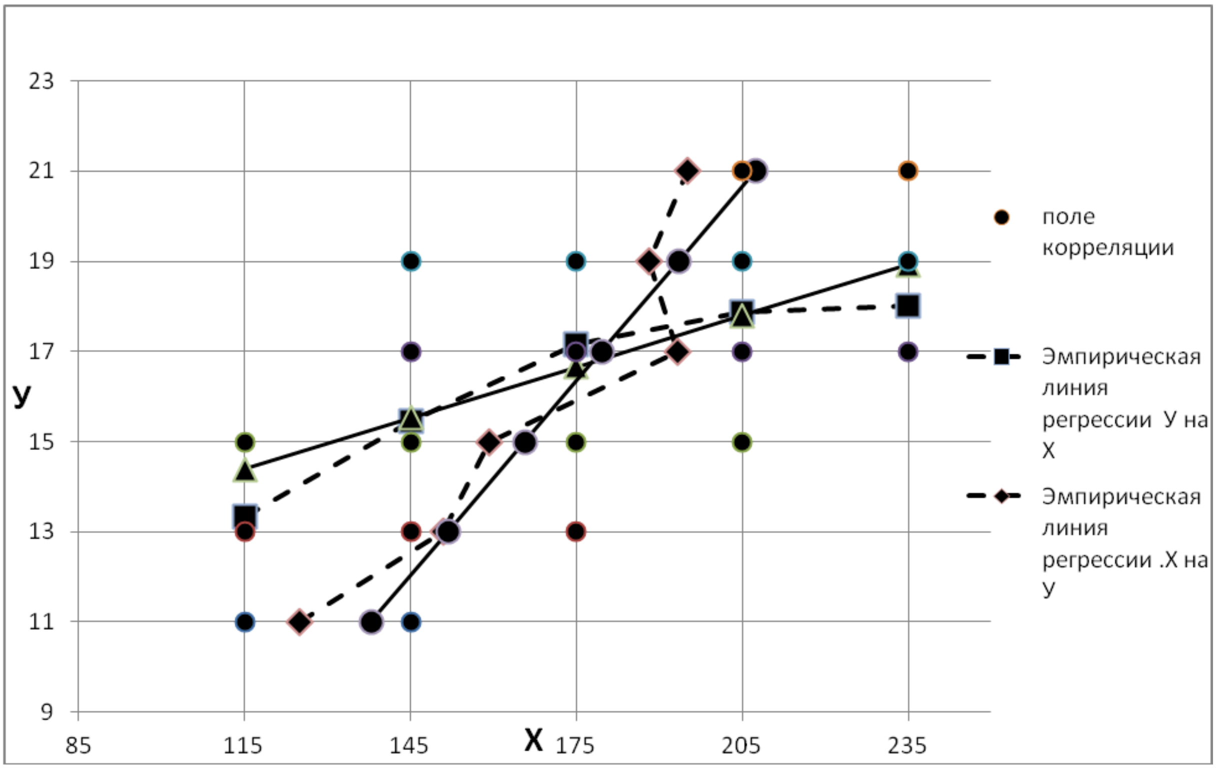

Рис. 8

Проверка нулевой гипотезы проводится на основании критерия Стьюдента. При справедливости нулевой гипотезы статистика:

$$\theta = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

будет иметь t-распределение Стьюдента с $k = n - 2$ степенями свободы.

В качестве альтернативной принимается гипотеза $H_1\colon \rho \neq 0$, т.е. коэффициент корреляции генеральной совокупности значимо отличен от нуля. Для такой альтернативной гипотезы критическая область будет двусторонней, симметричной относительно начала координат. Ее границы определяются из условия:

$$P\left(|\theta| > \theta_{кр.}\right) = \alpha.$$

Критическое значение определяется по таблице Стьюдента (приложение 4). При заданном уровне значимости α гипотеза H_0 отвергается, если $|\theta| > \theta_{кр.}$. В противном случае – нулевая гипотеза не отвергается.

Вычислим наблюдаемую или эмпирическую статистику критерия:

$$\theta_{набл.} = \frac{0{,}51 \cdot \sqrt{70-2}}{\sqrt{1-0{,}51^2}} \approx 4{,}89.$$

Для уровня значимости $\alpha = 0{,}05$ и числа степеней свободы $k = 70 - 2 = 68$ находим критическое значение статистики $\theta_{кр.} = \theta_{0{,}05;\,78} \approx 2{,}0$. Поскольку вычисленная $\theta_{набл.}$ статистика больше критического, то гипотеза H_0 отвергается, т.е. представленные эмпирические данные на заданном уровне значимости не противоречат альтернативной гипотезе о том, что коэффициент корреляции между объемом вложений в ценные бумаги и полученной от них прибыли значимо отличается от нуля.

Пример. 21. При исследовании корреляционной зависимости между объемом производства ξ и доходами от реализации продукции η получены следующие уравнения регрессии:

$$y = -0{,}5x + 7 \text{ и } x = -1{,}6y + 12.$$

Найти выборочные средние значения величин ξ и η, коэффициент корреляции, оценить тесноту и характер связи.

Решение. Воспользуемся свойством линий регрессии – прямые регрессии пересекаются в одной точке – общей средней. Следовательно, ее координаты должны удовлетворять обоим уравнениям. Тем самым, задача сводится к решению системы линейных уравнений вида:

$$\begin{cases} \bar{y} = -0{,}5\bar{x} + 7; \\ \bar{x} = -1{,}6\bar{y} + 12. \end{cases}$$

Решая последнюю систему, получим: $\bar{x} = 4,\ \bar{y} = 5.$ Так как коэффициенты регрессии отрицательны, то заданные уравнения описывают обратную корреляционную зависимость и коэффициент корреляции будет отрицательным:

$$r = -\sqrt{(-0{,}5)\cdot(-1{,}6)} = -\sqrt{0{,}8} \approx -0{,}89.$$

В силу того, что коэффициент корреляции по абсолютной величине больше 0,7, можно сказать, что связь между переменными является достаточно тесной.

§ 12. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ ПО ТЕМЕ «ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ»

20. При исследовании зависимости между вложениями в рекламу ξ (млн. руб.) и прибылью η (млн. руб.) среди 100 фирм получены следующие данные: $\bar{x}$ =2, $\bar{y}$ =10, b_{yx}=1,5, b_{xy}=0,06. Найти среднее значение прибыли при вложениях в рекламу, равных 1 млн. руб.

21. Исследование корреляционной зависимости удоя коров (ξ, л/день) от потребляемых ими концентратов (η, кг/день) показало, что средний надой от одной коровы в день составляет 10 л/день, а среднее потребление коровой концентратов равно 2 кг/день. Коэффициент регрессии $b_{y/x}$ составляет 2,5, а коэффициент корреляции равен 0,65. Записать уравнения регрессии и, используя соответствующее уравнение, найти среднее потребление концентратов в день при удое 12 л.

22. При исследовании корреляционной зависимости между удоем коров X и потреблением концентратов Y, получены следующие данные: $\bar{x} = 10$ л/день, $\bar{y} = 2$ кг/день, $s_x^2 = 6$, $s_y^2 = 0,5$, $\mu = 1,5$. Используя соответствующее уравнение регрессии, найти средний удой коров при потреблении 3 кг концентратов в день.

23. Известно, что выборочный коэффициент линейной корреляции равен 0,6. Уравнение регрессии η на ξ имеет вид $y = 0,6x - 2$. Найти $b_{x/y}$.

24. При исследовании корреляционной зависимости между величинами η и ξ получены следующие уравнения регрессии: $y = 0,8x - 1,3$ и $x = 0,2y + 1,1$. Найти средние значения признаков η и ξ и коэффициент линейной корреляции.

25. Распределение 50 компаний по ежемесячным затратам на рекламу (ξ, тыс.руб.) и объему выручки от продаж (η, млн. руб.) представлено в таблице:

Затраты на рекламу, тыс. руб. (ξ)	*Объем выручки от продаж, млн.руб. (η)*						*Итого*
	28 – 32	32 – 36	36 – 40	40 – 44	44 – 48	48 – 52	
10 – 15	2	3	1				6
15 – 20		1	6	3			10
20 – 25		1	3	8	4		16
25 – 30			1	2	6	2	11
30 – 35				1	2	4	7
Итого	2	5	11	14	12	6	**50**

Необходимо:

1. Вычислить групповые средние $\bar{x}_i$ и $\bar{y}_j$, построить поле корреляции и эмпирические линии регрессии.

2. Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

б) используя соответствующее уравнение регрессии, оценить полученную прибыль при объеме вложений в ценные бумаги 150 млн. руб.;

в) вычислить коэффициент линейной корреляции, сделать вывод о тесноте и направлении связи между переменными ξ и η;

г) на уровне значимости α=0,05 с помощью критерия Стьюдента оценить значимость коэффициента корреляции.

26. При исследовании корреляционной зависимости между объемом производства X и доходами от реализации продукции Y получены следующие уравнения регрессии: $y = 0{,}3x + 120$ и $x = 1{,}6y - 88$. Найти среднее значение величины Y.

§ 13. ВАРИАНТЫ РАСЧЁТНО-АНАЛИТИЧЕСКОЙ РАБОТЫ И ТРЕБОВАНИЯ К ЕЁ ОФОРМЛЕНИЮ

При выполнении расчётно-аналитической работы необходимо придерживаться правил, приведённых ниже.

1. Индивидуальный номер варианта соответствует последней цифре номера личного дела студента, который совпадает с номером зачётной книжки и студенческого билета.

   Расчётно-аналитическая работа не рассматривается, если её вариант не совпадает с последней цифрой номера личного дела студента.

2. Следует правильно оформить титульный лист в соответствии с предлагаемым ниже образцом.
3. Необходимо приводить условия каждой из решаемых задач прежде, чем начинается её решение.
4. Все этапы решения должны содержать развёрнутые объяснения и описание вводимых обозначений. Используемые формулы и теоремы должны записываться с необходимыми пояснениями. Окончательный ответ следует выделить и сформулировать словесно.
5. Все расчёты нужно проводить с учетом правил приближённых вычислений. Учитывая, что используемые при решении задач таблицы четырехзначные, все промежуточные вычисления следует проводить с четырьмя верными знаками после запятой, а окончательный ответ дать с тремя верными знаками, правильно округлив полученный до этого результат.
6. В конце работы указывается список использованной литературы, ставится дата окончания работы и подпись.
7. Рекомендуется справа оставлять поля для возможных замечаний проверяющего.
8. Работа может быть представлена в любом доступном формате и прикрепляется на портал для проверки, если задана контрольная точка.
9. Расчётно-аналитическая работа в распечатанном виде обязательно предъявляется на экзамене. В случае успешной сдачи экзамена работа остается у экзаменатора.

**ОБРАЗЕЦ ТИТУЛЬНОГО ЛИСТА
РАСЧЕТНО-АНАЛИТИЧЕСКОЙ РАБОТЫ.**

Федеральное государственное образовательное бюджетное учреждение
высшего образования
«Финансовый университет при Правительстве Российской Федерации»
(Финуниверситет)

Департамент анализа данных, принятия решений и финансовых технологий

**Расчетно-аналитическая работа
по дисциплине
«Анализ данных»**

Вариант № ____ (номер личного дела____________)

Выполнил(а) студент(ка) группы ______(название группы)

Направление подготовки______38.03.01 «Экономика»__

_________________________________ (Ф.И.О.)

Преподаватель _____________(Ф.И.О. преподавателя).

Москва 2019

ВАРИАНТ 1

(для студентов, номера личных дел которых оканчиваются цифрой 1)

1. Вероятность сдачи студентом контрольной работы в срок равна 0,8. Найти вероятность того, что из 60 студентов вовремя сдадут контрольную работу:

а) 45 студентов;

б) не менее половины студентов;

в) не менее 40, но не более 50 студентов.

2. По наблюдениям за температурой воздуха в сентябре этого года в данной местности установлено, что средняя температура воздуха составила 15ºС, а среднее квадратическое отклонение равно 5ºС, Оценить вероятность того, что в сентябре следующего года средняя температура воздуха будет:

а) не более 25ºС;

б) более 20ºС.

в) будет отличаться от средней температуры этого года не более чем на 7ºС (по абсолютной величине);

г) будет отличаться от средней температуры этого года не менее чем на 8ºС (по абсолютной величине);

3. Известно, что месячная доходность некоторой ценной бумаги есть нормально распределенная случайная величина ξ (%). Найти ее математическое ожидание и среднее квадратическое отклонение, если известно, что $P(\xi < 1) = 0{,}1$ и $P(\xi \geq 5) = 0{,}5$.

Построить графики функции распределения и функции плотности распределения этой случайной величины.

Вычислить вероятность того, что в следующем месяце доходность ценной бумаги будет:

а) не более 4%;

б) не менее 8%;

в) от 3% до 7%.

4. С целью изучения соблюдения трудовой дисциплины было обследовано 100 предприятий из 500 (выборка бесповторная). Получены следующие данные о количестве зарегистрированных нарушений:

Количество нарушений	Менее 3	3-5	5-7	7-9	9-11	Более 11	Итого
Число предприятий	10	17	27	23	15	8	100

Найти:

1) вероятность того, что среднее количество нарушений на всех предприятиях отличается от их среднего количества в выборке не более, чем на одно;

2) границы, в которых с вероятностью 0,98 заключена доля предприятий, где количество нарушений превышает 9.

3) объем бесповторной выборки, при котором те же границы для среднего количества нарушений, что и в п. 1 можно гарантировать с вероятностью 0,95.

5. С целью изучения миграции населения в данной области было проведено выборочное обследование 70 мелких населенных пунктов из 350 имеющихся в области (выборка бесповторная). Получены следующие данные о количестве зарегистрированных мигрантов:

7	7	5	2	7	4	10	4	10	0
2	3	7	4	3	8	3	10	9	9
9	1	1	2	6	10	0	7	3	5
2	8	4	2	10	5	1	1	4	7
6	4	6	10	5	5	7	2	8	5
7	6	9	1	6	7	8	6	8	5
9	8	6	6	6	7	4	7	7	4

Составить интервальный вариационный ряд. Записать эмпирическую функцию распределения и построить ее график. На одном чертеже изобразить гистограмму и полигон частот.

По сгруппированным данным вычислить выборочные числовые характеристики: среднее арифметическое, исправленную выборочную дисперсию, среднее квадратичное отклонение, коэффициент вариации, асимметрию, эксцесс, моду и медиану.

Заменив параметры генеральной совокупности соответственно их наилучшими выборочными числовыми характеристиками, используя χ^2-критерий Пирсона, на уровне значимости α=0,05 проверить две гипотезы о том, что изучаемая случайная величина ξ – число мигрантов в данном населенном пункте – распределена:

а) по нормальному закону распределения;

б) по равномерному закону распределения.

Построить чертёж, на котором изображена гистограмма эмпирического распределения, соответствующие графики равномерного и нормального распределений.

6. С целью изучения зависимости количества времени использования клиентом мобильной связи в течение месяца ξ (мин.) и стоимости минуты разговора η (руб.) произведено обследование 100 абонентов, пользующихся различными тарифными планами, и получены следующие данные:

ξ \ η	Менее 1	1-1,5	1,5-2	2-2,5	2,5-3	Более 3	Итого
Менее 200				3	9	3	15
200-400				5	8	7	20
400-600			4	13	9	3	29
600-800		2	6	8	2		18
Более 800	6	5	6	1			18
Итого	6	7	16	30	28	13	100

Необходимо:

1. Вычислить групповые средние $\overline{x}_i$ и $\overline{y}_j$, построить эмпирические линии регрессии.

2. Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

б) вычислить коэффициент корреляции Пирсона; на уровне значимости α= 0,05 оценить его значимость и сделать вывод о тесноте и направлении связи между переменными ξ и η;

в) используя соответствующее уравнение регрессии, оценить время использования мобильной связи при стоимости минуты разговора 2,25 руб.

ВАРИАНТ 2

(для студентов, номера личных дел которых оканчиваются цифрой 2)

1. В отделении Сбербанка микрорайона пользуются банкоматом 20% населения из близлежащих домов. Какова вероятность того, что из 500 наудачу выбранных жителей микрорайона в этом отделении Сбербанка пользуются банкоматом:

а) 90 человек;

б) от 80 до 130 человек;

б) более 120 человек?

2. Всхожесть хранящегося на складе зерна в среднем составляет 80%, а среднее квадратическое отклонение 6%. Оценить вероятность того, что в выбранной партии зерна всхожесть составит:

а) не менее 85%;

б) не более 90%;

в) будет отличаться от средней не более чем на 8%;

г) будет отличаться от средней не менее чем на 10%.

3. Случайная величина ξ имеет нормальный закон распределения с параметрами a и σ^2. Найти параметр σ, если известно, что $M(\xi)$=5 и $P(2<\xi<8)=0{,}9973$. Вычислить вероятность того, что значение случайной величины ξ окажется меньше 0.

Построить графики функции распределения и функции плотности распределения этой случайной величины.

4. Из 1560 сотрудников предприятия по схеме собственно случайной бесповторной выборки отобрано 100 человек для получения статистических данных о пребывании на больничном листе в течение года. Полученные данные представлены в таблице:

Количество дней пребывания на больничном листе	Менее 3	3–5	5–7	7–9	9–11	Более 11	Итого
Число сотрудников	6	13	24	39	8	10	100

Найти:

а) вероятность того, что среднее число дней пребывания на больничном листе среди сотрудников предприятия отличается от их среднего числа в выборке не более чем на 1 день (по абсолютной величине);

б) границы, в которых с вероятностью 0,95 заключена доля всех сотрудников, пребывающих на больничном листе не более 7 дней;

в) объем бесповторной выборки, при котором те же границы для доли, (см. п. б)), можно гарантировать с вероятностью 0,98.

5. С целью изучения размера потребительских кредитов, выданных банком в одном из крупных магазинов электронной техники в течении последнего месяца по схеме собственно-случайной бесповторной выборки было отобрано 180 кредитов из 2500 выданных. Величины сумм выданных кредитов (тыс. руб.) представлены в таблице:

22,9	26,6	18,0	25,2	28,9	30,3	21,1	13,5	15,7	22,2
18,6	28,8	11,5	26,7	31,6	14,1	26,7	22,2	19,9	23,4
16,0	17,9	17,0	20,3	10,5	26,8	13,9	18,1	19,6	12,7
20,7	17,8	19,5	24,4	21,8	23,3	18,6	24,1	19,6	20,8
15,8	14,0	20,5	18,2	17,8	20,7	21,9	28,0	17,5	11,2
12,2	24,7	14,9	19,3	23,6	22,3	20,1	19,1	21,9	25,2
22,2	18,0	16,3	18,3	18,6	13,5	28,0	15,2	22,1	24,7
20,1	14,0	17,3	17,6	18,9	22,4	20,9	15,1	11,9	21,8
23,4	18,2	21,0	22,7	23,2	19,9	26,1	21,3	21,2	16,1
27,6	17,5	18,1	13,0	23,9	11,2	22,5	19,5	19,2	24,2
29,7	22,7	12,7	26,4	16,8	14,7	21,3	18,5	22,3	15,3
14,0	23,1	25,8	27,9	17,5	24,9	25,6	32,4	17,9	19,7
11,9	17,6	15,0	19,0	22,1	14,0	27,5	18,6	19,5	25,5
19,5	25,3	27,9	24,9	15,5	13,8	24,2	23,8	25,8	18,9
8,3	24,6	18,7	24,2	16,3	18,9	22,4	15,6	25,6	16,6
19,6	20,0	20,2	9,9	22,0	19,2	14,5	12,6	13,0	20,1
22,7	20,7	20,2	12,9	21,1	19,0	20,2	28,0	20,2	21,8
14,8	17,3	17,4	14,1	13,8	19,2	17,0	22,0	17,1	17,2

Составить интервальный вариационный ряд. Записать эмпирическую функцию распределения и построить ее график. На одном чертеже изобразить гистограмму и полигон частот.

По сгруппированным данным вычислить выборочные числовые характеристики: среднее арифметическое, исправленную выборочную дисперсию, среднее квадратичное отклонение, коэффициент вариации, асимметрию, эксцесс, моду и медиану.

Заменив параметры генеральной совокупности соответственно их наилучшими выборочными числовыми характеристиками и используя χ^2-критерий Пирсона, на уровне значимости α=0,05 проверить две гипотезы о том, что изучаемая случайная величина ξ – величина выданных кредитов – распределена:

а) по нормальному закону распределения;

б) по равномерному закону распределения.

Построить на чертеже, на котором изображена гистограмма эмпирического распределения и соответствующие графики равномерного и нормального распределений.

6. В таблице приведено распределение 120 коров по дневному надою ξ (в кг) и жирности молока η (в %):

ξ / η	Менее 7	7 -10	10 - 13	13 - 16	Более 16	Итого
Менее 3,2				8		8
3,2 – 3,6			2	16	8	26
3,6 – 4,0		4	16	10	2	32
4,0 – 4,4	2	6	10	2		20
Более 4,4	8	6	20			34
Итого	10	16	48	36	10	120

1) Вычислить групповые средние $\overline{x}_i$ $и$ $\overline{y}_j$ и построить эмпирические линии регрессии;

2) Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать содержательную интерпретацию полученных уравнений;

б) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0{,}05$ оценить его значимость и сделать вывод о тесноте и направлении связи между переменными ξ и η;

в) используя соответствующее уравнение регрессии, оценить средний процент жирности молока для коров, дневной удой которых составляет 15 кг.

ВАРИАНТ 3

(для студентов, номера личных дел которых оканчиваются цифрой 3)

1. Торговая фирма продала 600 телевизоров. Если телевизор оказывается неисправным, фирма забирает его у потребителя, отправляет поставщику и несет при этом расход 100 тыс. рублей. Какова вероятность того, что расходы составят более десяти миллионов рублей, если в среднем с дефектами оказывается каждый двенадцатый телевизор?

2. Уровень воды в реке – это случайная величина со средним значением 2,5 м и стандартным отклонением 20 см. Оценить вероятность того, что в наудачу выбранный день:

а) уровень превысит 3 м;

б) уровень не превысит 275 см;

в) будет отличаться от среднего уровня более чем на 40 см;

г) окажется в пределах от 2м 20см до 2м 80см.

3. Известно, что время непрерывной работы электрической лампы есть случайная величина ξ (час.), имеющая показательный закон распределения. Найти математическое ожидание и среднее квадратическое отклонение этой случайной величины, если известно, что вероятность непрерывной работы лампы не менее 800 час составляет 0,2.

Построить графики функции распределения и функции плотности распределения этой случайной величины.

Вычислить вероятность того, что выбранная случайным образом лампа непрерывно проработает:

а) не более 600 час;

б) не менее 700 час;

в) от 30 до 40 суток.

4. В некоторой области по схеме собственно случайной бесповторной выборки было обследовано 80 предприятий малого бизнеса из 2500 с целью изучения объема привлечённых инвестиций. Получены следующие данные:

Объём привлечённых инвестиций, тыс.руб.	Менее 600	600–700	700–800	800–900	900–1000	Более 1000	Итого
Число предприятий	12	19	23	18	5	3	80

Найти:

а) вероятность того, что средний объем привлечённых инвестиций во всех предприятиях малого бизнеса в области отличается от среднего объема привлечённых инвестиций, полученного в выборке, не более чем на 15 тыс.руб (по абсолютной величине);

б) границы, в которых с вероятностью 0,98 заключена доля предприятий, с объемом инвестиций от 600 до 900 тыс.руб;

в) объем бесповторной выборки, при котором те же границы для среднего объема инвестиций (см. п. а)) можно гарантировать с вероятностью 0,95.

5. С целью определения средней суммы вкладов на 1 января текущего года в сберегательном банке, имеющем 2000 вкладчиков, по схеме собственно-случайной выборки с бесповторным отбором членов проведено обследование 200 лицевых счетов. Распределение вкладов по их величине (тыс. руб.) представлено в таблице:

612	442	498	284	667	563	709	388	518	717
218	600	605	131	547	517	448	818	732	842
501	385	238	682	400	498	305	610	463	618
537	453	546	723	190	608	607	620	117	705
562	212	520	414	316	408	405	355	457	569
367	429	254	568	413	572	423	755	154	588
594	473	340	335	566	402	401	502	756	558
792	565	474	526	502	408	674	828	483	465
596	670	502	601	452	523	741	261	327	556
541	496	141	274	394	555	409	511	644	560
549	763	739	455	475	287	522	743	535	630
494	562	488	562	656	559	540	592	591	348
498	495	457	644	379	877	398	272	363	597
231	539	667	583	369	492	559	662	239	532
574	568	621	663	223	714	649	476	619	428
494	567	536	359	502	511	389	621	573	305
520	561	634	609	563	359	343	702	489	136
725	495	507	627	775	489	419	430	598	511
661	593	386	643	182	366	611	464	665	427
389	779	761	644	607	536	706	694	462	354

Составить интервальный вариационный ряд. Записать эмпирическую функцию распределения и построить ее график. На одном чертеже изобразить гистограмму и полигон частот.

По сгруппированным данным вычислить выборочные числовые характеристики: среднее арифметическое, исправленную выборочную дисперсию, среднее квадратичное отклонение, коэффициент вариации, асимметрию, эксцесс, моду и медиану.

Заменив параметры генеральной совокупности соответственно их наилучшими выборочными числовыми характеристиками и используя χ^2-

критерий Пирсона, на уровне значимости α=0,05 проверить две гипотезы о том, что изучаемая случайная величина ξ – величина вклада – распределена:

а) по нормальному закону распределения;

б) по равномерному закону распределения.

Построить чертёж, на котором изображена гистограмма эмпирического распределения и соответствующие графики равномерного и нормального распределений.

6. Распределение 100 средних фермерских хозяйств по числу наемных рабочих ξ (чел.) и их средней месячной заработной плате на 1 человека η (тыс. руб.) представлено в таблице:

$\eta \backslash \xi$	2 – 4	4 – 6	6 – 8	8 – 10	10 – 12	Итого
20–25			6	8	4	18
25–30		2	10	2	2	16
30–35	2	6	8	2		18
35–40	4	12	10	2		28
40–45	10	6	4			20
Итого	16	26	38	14	6	100

1) Вычислить групповые средние $\overline{x}_i$ *и* $\overline{y}_j$ и построить эмпирические линии регрессии;

2) Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

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

в) используя соответствующее уравнение регрессии, оценить среднюю месячную заработную плату одного рабочего в хозяйстве, в котором работают 7 наемных рабочих.

ВАРИАНТ 4

(для студентов, номера личных дел которых оканчиваются цифрой 4)

1. В городе в среднем 10% заключенных браков в течение года заканчиваются разводом. Какова вероятность того, что из 200 случайно отобранных пар, заключивших брак, в течение года не разведутся:

а) 170 пар;

б) не менее 185 пар.

2. Говорят, что вероятность того, что желание, загаданное на Новый год, сбудется равна 0,7. Если это правда, то оцените вероятность того, что из 200 загаданных желаний число сбывшихся будет

а) не менее 130;

б) заключено в интервале от 120 до 160.

Оценить, сколько желаний надо загадать, чтобы с вероятностью не меньшей 0,95, исполнилось бы не менее 40% желаний.

3. Плотность вероятности случайной величины ξ имеет вид:

$$\varphi(x)=\begin{cases} 0 & \text{при} \quad x\le 0, \\ ax & \text{при} \quad 0<x\le 4, \\ 0 & \text{при} \quad x>4. \end{cases}$$

Найти:

а) параметр распределения a;

б) функцию распределения $F(x)$;

в) математическое ожидание $M(\xi)$ и дисперсию $D(\xi)$;

г) вероятность $P(0<\xi<1)$. Построить графики функций $\varphi(x)$ и $F(x)$.

4. В результате выборочного обследования российских автомобилей, которые обслуживаются в автосервисе по гарантии, по схеме собственно-случайной бесповторной выборки из 280 автомобилей были отобраны 60. Полученные данные о пробеге автомобилей с момента покупки до первого гарантийного ремонта представлены в таблице:

Пробег, тыс. км	Менее 1	1–2	2–3	3–4	4–5	5–6	Более 6	Итого
Число автомобилей	3	5	9	16	13	8	6	60

Найти:

а) вероятность того, что средний пробег всех автомобилей отличается от среднего пробега автомобилей в выборке не более, чем на 400 км (по абсолютной величине);

б) границы, в которых с вероятностью 0,95 заключена доля автомобилей, пробег которых составляет менее 3 тыс. км.;

в) объем бесповторной выборки, при котором те же границы для доли (см. п. б)), можно гарантировать с вероятностью 0,9876.

5. С целью определения средней величины транспортных затрат (тыс. руб.) на доставку одной тонны продукции предприятий пищевой промышленности к потребителям в некотором крупном мегаполисе, имеющем 2570 предприятий, по схеме собственно-случайной бесповторной выборки проведено обследование 240 предприятий. Распределение транспортных затрат (тыс. руб.) представлено в таблице:

10,3	8,8	6,8	14,0	8,8	13,2	8,2	9,5	9,9	14,0
13,2	14,4	11,7	10,7	6,8	11,5	10,8	8,2	8,2	6,2
5,3	11,7	4,0	6,2	13,6	18,1	7,6	10,7	13,0	14,8
10,0	11,2	6,2	9,3	11,6	6,6	10,1	6,5	9,1	11,9
10,2	9,7	11,0	4,3	8,6	12,9	15,9	9,7	12,7	6,0
9,6	14,0	7,9	10,6	8,8	11,9	15,6	8,3	6,8	3,4
5,1	11,5	12,8	12,6	9,8	12,0	7,7	6,7	9,6	11,8

10,5	10,7	10,3	6,8	13,0	7,5	9,1	11,0	8,0	10,0
9,5	4,6	6,6	9,5	10,2	9,5	14,7	16,3	17,8	9,5
10,0	7,6	11,9	10,6	3,8	10,9	7,9	14,4	8,0	9,7
12,6	14,4	8,2	13,9	6,2	9,9	7,1	12,1	7,6	9,0
6,4	10,9	8,4	13,5	8,3	4,5	5,9	15,6	13,7	12,6
8,4	11,3	12,8	12,8	7,7	14,0	8,9	9,7	9,8	14,1
7,0	8,2	8,4	13,9	7,9	11,7	8,5	9,7	2,6	11,5
6,6	8,4	0,6	12,2	12,1	12,4	11,3	11,7	6,5	12,9
10,6	8,8	12,0	11,0	9,4	7,0	13,0	14,4	9,3	13,6
12,7	5,7	5,8	9,5	11,0	11,8	9,9	7,9	12,4	9,0
10,6	10,9	9,8	10,9	10,9	5,7	11,6	8,7	12,5	7,0
13,6	10,3	11,1	13,5	12,0	9,1	9,3	7,3	15,3	12,1
3,7	10,7	9,4	7,4	14,5	9,5	10,5	9,1	8,5	12,8
11,8	1,9	13,4	12,9	11,2	9,4	15,0	12,7	10,5	10,0
16,1	11,5	11,1	10,4	4,8	13,0	7,7	9,0	11,1	10,0
17,0	9,6	8,7	9,4	15,6	9,6	9,3	9,4	13,9	12,1
8,2	2,0	12,5	10,0	11,2	8,2	5,8	11,3	8,2	9,4

Составить интервальный вариационный ряд. Записать эмпирическую функцию распределения и построить ее график. На одном чертеже изобразить гистограмму и полигон частот.

По сгруппированным данным вычислить выборочные числовые характеристики: среднее арифметическое, исправленную выборочную дисперсию, среднее квадратичное отклонение, коэффициент вариации, асимметрию, эксцесс, моду и медиану.

Заменив параметры генеральной совокупности соответственно их наилучшими выборочными числовыми характеристиками и используя χ^2-критерий Пирсона, на уровне значимости α=0,05 проверить две гипотезы о том, что изучаемая случайная величина ξ – величина транспортных затрат – распределена:

а) по нормальному закону распределения;

б) по равномерному закону распределения.

Построить чртёж, на котором изображена гистограмма эмпирического распределения и соответствующие графики равномерного и нормального распределений.

6. Распределение 60 предприятий по затратам рабочего времени ξ (тыс. чел. дн.) и выпуску продукции η (млн. руб.) представлено в таблице:

η \ ξ	30–40	40–50	50–60	60–70	70–80	Итого:
10–25	1	3	2			6
25–40	3	6	4	1		14
40–55		3	7	6	1	17
55–70		1	6	4	4	15
70–85			2	5	1	8
Итого:	4	13	21	16	6	60

1) Вычислить групповые средние $\overline{x}_i$ *и* $\overline{y}_j$ и построить эмпирические линии регрессии;

2) Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

б) вычислить коэффициент корреляции; на уровне значимости $\alpha = 0{,}05$ оценить его значимость и сделать вывод о тесноте и направлении связи между переменными ξ и η;

в) используя соответствующее уравнение регрессии, оценить средний выпуск продукции предприятия с затратами рабочего времени 55 тыс. чел. дн.

ВАРИАНТ 5

(для студентов, номера личных дел которых оканчиваются цифрой 5)

1.После окончания занятий в среднем каждый десятый студент занимается в читальном зале.

Найти вероятность того, что из 200 студентов будут заниматься в читальном зале:

а) 15 студентов;

б) не менее 15, но не более 30 студентов;

в) сколько посадочных мест нужно иметь, чтобы с вероятностью 0,9545 их хватало всем желающим заниматься в читальном зале студентам

2. Дневная выручка магазина шаговой доступности является случайной величиной, распределенной по нормальному закону со средним значением 25000 руб. и средним квадратическим отклонением 3000 руб.

1) С помощью неравенства Чебышева оценить вероятность того, что дневная выручка магазина шаговой доступности будет находится в пределах от 22000 до 28000 руб.

2) Ту же вероятность найти, используя связь нормального закона распределения с функцией Лапласа.

3. Функция распределения непрерывной случайной величины ξ имеет вид:

$$F(x)=\begin{cases} 0 & \text{при} \quad x \le 2, \\ a(x-2) & \text{при} \quad 2 < x \le 2{,}5, \\ 1 & \text{при} \quad x > 2{,}5. \end{cases}$$

Найти: 1) параметр a;

2) плотность вероятности $\varphi(x)$;

3) математическое ожидание $M(\xi)$ и дисперсию $D(\xi)$.

Построить графики функций $\varphi(x)$ и $F(x)$.

4. В филиале заочного вуза обучается 2000 студентов. Для изучения стажа работы студентов по специальности по схеме собственно случайной бесповторной выборки отобрано 100 студентов. Полученные данные о стаже работы студентов по специальности представлены в таблице:

Стаж работы по специальности, лет	Менее 2	2–4	4–6	6–8	8–10	10–12	Более 12	Итого
Количество студентов	10	19	24	27	12	5	3	100

Найти:

а) вероятность того, что доля всех студентов филиала, имеющих стаж работы менее 6 лет, отличается от выборочной доли таких студентов не более чем на 5% (по абсолютной величине);

б) границы, в которых с вероятностью 0,997 заключен средний стаж работы по специальности всех студентов филиала;

в) объем бесповторной выборки, при котором те же границы для среднего стажа работы по специальности (см. п. б)) можно гарантировать с вероятностью 0,9898.

5. По схеме собственно-случайной бесповторной выборки проведено 10%-ное обследование строительных организаций региона по недельному объему выполненных строительных работ (тыс. руб.). Предполагая, что в регионе функционируют 1300 строительных организаций, получены следующие данные:

748	449	713	602	775	661	1047	676	1008	488
612	641	761	660	642	794	636	924	859	866
839	573	510	597	735	1035	435	759	645	695
597	795	671	596	922	694	556	572	668	776
729	656	738	941	702	707	479	610	783	698
824	877	572	887	649	984	668	857	616	498
682	716	749	706	667	865	896	697	519	841
838	838	711	609	740	433	714	940	848	561
609	837	715	766	451	603	639	673	613	821
784	665	534	751	580	748	753	629	686	724
728	643	701	617	687	540	834	867	804	756
610	712	828	779	739	686	556	824	755	650
833	882	521	509	849	870	825	891	749	853

Составить интервальный вариационный ряд. Записать эмпирическую функцию распределения и построить ее график. На одном чертеже изобразить гистограмму и полигон частот.

По сгруппированным данным вычислить выборочные числовые характеристики: среднее арифметическое, исправленную выборочную дисперсию, среднее квадратичное отклонение, коэффициент вариации, асимметрию, эксцесс, моду и медиану.

Заменив параметры генеральной совокупности соответственно их наилучшими выборочными числовыми характеристиками и используя χ^2-критерий Пирсона, на уровне значимости α=0,05 проверить две гипотезы о том, что изучаемая случайная величина ξ – величина транспортных затрат – распределена:

а) по нормальному закону распределения;

б) по равномерному закону распределения.

Построить чертёж, на котором изображена гистограмма эмпирического распределения и соответствующие графики равномерного и нормального распределений.

6. Распределение 50 городов по численности населения ξ (тыс. чел.) и среднемесячному доходу на одного человека η (тыс. руб.) представлено в таблице:

η \ ξ	3–4	4–5	5–6	6–7	7–8	Более 8	Итого
30–50	1	1	3				5
50–70		2	5	1			8
70–90		1	1	6	2	2	12
90–110			4	9			13
110–130			2	2	5		9
Более 130					2	1	3
Итого:	1	4	15	18	9	3	50

Необходимо:

1) Вычислить групповые средние $\overline{x}_i$ и $\overline{y}_j$, построить эмпирические линии регрессии;

2) Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

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

в) используя соответствующее уравнение регрессии, оценить средний доход на одного человека в городе с населением 100 тыс. человек.

ВАРИАНТ 6

(для студентов, номера личных дел которых оканчиваются цифрой 6)

1. На почту поступило 8000 писем. Вероятность того, что на случайно взятом конверте отсутствует почтовый индекс, равна 0,0005. Найти вероятность того, что почтовый индекс отсутствует: а) на трех конвертах; б) не менее чем на трех конвертах.

2. Вероятность поражения цели при одном выстреле равна 0,4. Было произведено 600 выстрелов. Найти:

а) границы, в которых с вероятностью 0,9949 будет заключено число попаданий в цель;

б) число выстрелов, которые надо произвести по мишени, чтобы с вероятностью 0,9949 ожидать, что отклонение относительной частоты от вероятности попадания при одном выстреле будет по модулю меньше 0,05.

3. Случайная величина ξ подчинена нормальному закону распределения с нулевым математическим ожиданием. Вероятность попадания этой случайной величины в интервале от –2 до 2 равна 0,5705. Найти среднее квадратическое отклонение и плотность распределения этой случайной величины.

4. На предприятии работает 2000 сотрудников. Для изучения стажа работы сотрудников на этом предприятии по схеме собственно случайной бесповторной выборки отобрано 400 человек. Полученные данные о стаже работы представлены в таблице:

Стаж работы, годы	До 5	5–10	10–15	15–20	Свыше 20	Итого
Число работников	32	56	92	120	100	400

Найти:

а) вероятность того, что средний стаж работы отличается от среднего стажа в выборке не более чем на 2 года (по абсолютной величине);

б) границы, в которых с вероятностью 0,95 заключена доля работников, стаж которых менее 7 лет;

в) объем повторной выборки, при которой те же границы для доли работников (см. п. б)) можно гарантировать с вероятностью 0,9876; дать ответ на тот же вопрос, если никаких предварительных данных о рассматриваемой доле нет.

5. С целью определения средней величины месячной заработанной платы работников торговой сферы в некотором крупном районе города, по схеме собственно-случайной бесповторной выборки было отобрано 150 работников из 1300. Распределение месячной заработной платы (тыс. руб.) представлено в таблице:

18,3	23,3	20,2	29,9	33,5	22,2	17,3	23,7	21,7	21,3
29,8	25,9	28,7	32,1	25,4	24,8	31,8	24,8	19,0	27,0
18,1	21,8	20,9	21,4	19,8	36,6	32,6	20,5	28,6	31,4
30,1	31,2	31,7	23,2	25,3	22,3	11,1	36,8	25,1	27,2
25,5	34,0	4,7	18,7	30,2	26,4	20,3	13,3	20,1	22,6
33,0	29,8	24,8	27,7	30,7	34,3	20,7	34,0	18,6	34,5
28,6	32,2	21,7	28,8	33,2	30,6	22,4	29,7	33,6	22,3
22,5	16,3	28,2	21,4	30,6	33,4	20,9	24,2	29,7	43,1
16,0	18,3	22,1	25,7	21,4	16,7	24,3	17,0	35,8	23,7
17,7	27,4	21,7	25,9	29,8	29,7	33,6	12,0	7,0	23,6
20,0	37,6	41,7	29,7	29,9	25,8	29,4	26,9	15,8	27,2
32,6	26,9	15,3	21,9	21,9	23,7	20,5	25,5	22,5	22,3
30,7	21,9	23,1	31,6	18,8	35,3	21,8	20,6	24,3	25,6
11,4	35,4	30,1	22,7	25,3	32,4	28,3	21,7	24,7	25,6
27,9	18,8	32,6	18,7	27,7	26,3	34,2	23,7	25,0	30,2

Составить интервальный вариационный ряд. Записать эмпирическую функцию распределения и построить ее график. На одном чертеже изобразить гистограмму и полигон частот.

По сгруппированным данным вычислить выборочные числовые характеристики: среднее арифметическое, исправленную выборочную дисперсию, среднее квадратичное отклонение, коэффициент вариации, асимметрию, эксцесс, моду и медиану.

Заменив параметры генеральной совокупности соответственно их наилучшими выборочными числовыми характеристиками и используя χ^2-критерий Пирсона, на уровне значимости α=0,05 проверить две гипотезы о том, что изучаемая случайная величина ξ – величина транспортных затрат – распределена:

а) по нормальному закону распределения;

б) по равномерному закону распределения.

6. Распределение 50 квартир в некотором городе по их стоимости η млн руб и площади ξ кв.м задано в таблице:

ξ \ η	3–4	4–5	5–6	6–7	7–8	Более 8	Итого
30–50	1	1	3				5
50–70		2	5	1			8
70–90		1	1	6	2	2	12
90–110			4	9			13
110–130			2	2	5		9
Более 130					2	1	3
Итого:	1	4	15	18	9	3	50

1) Вычислить групповые средние $\bar{x}_i$ и $\bar{y}_j$ и построить эмпирические линии регрессии;

2) Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

б) вычислить коэффициент корреляции; на уровне значимости $\alpha = 0{,}05$ оценить его значимость и сделать вывод о тесноте и направлении связи между переменными ξ и η;

в) используя соответствующее уравнение регрессии, оценить среднюю стоимость квартиры с площадью 52 кв.м.

ВАРИАНТ 7

(для студентов, номера личных дел которых оканчиваются цифрой 7)

1. При установившемся технологическом процессе в день в среднем происходит 10 обрывов нити на 100 веретенах. Определить вероятность того, что на 800 веретенах произойдет:

а) ровно 78 обрывов нити;

б) обрыв нити произойдет не более чем на 100 веретенах.

2. Опыт работы страховой компании показывает, что страховой случай приходится примерно на каждый десятый договор. Оценить с помощью неравенства Чебышева необходимое количество договоров, которые необходимо заключить, чтобы с вероятностью 0,9 можно было утверждать, что доля страховых случаев отклонится от 0,1 не более, чем на 0,01 (по абсолютной величине).

3. Диаметр выпускаемой детали является нормально распределенной случайной величиной с математическим ожиданием a=5 см и средним квадратическим отклонением σ=0,02 см. Найти вероятность того, что из двух проверенных деталей, диаметр хотя бы одной отклоняется от математического ожидания не более, чем на 0,04 см (по абсолютной величине).

4. В результате выборочного обследования 100 предприятий региона из 500 по схеме собственно случайной бесповторной выборки получено следующее распределение снижения затрат на производство продукции в процентах к предыдущему году:

Процент снижения затрат (%)	4–6	6–8	8–10	10–12	12–14	14–16	Итого
Число предприятий	6	20	31	24	13	6	100

Найти:

а) границы, в которых с вероятностью 0,907 будет находиться средний процент снижения затрат на всех 500 предприятиях;

б) вероятность того, что доля всех предприятий, затраты которых снижены не менее, чем на 10%, отличается от доли таких предприятий в выборке не более, чем на 0,04 (по абсолютной величине);

в) объем бесповторной выборки, при котором те же границы для среднего процента сниженные затрат (см. п. а)) можно гарантировать с вероятностью 0,9876.

5. В некотором городе по схеме собственно случайной бесповторной выборки было обследовано 180 магазинов розничной торговли из 2500 с целью изучения месячного объема розничного товарооборота. Распределение месячного объема розничного товарооборота (тыс. руб.) представлено в таблице:

284	492	443	351	698	423	403	418	881	485
697	693	656	679	517	513	458	554	303	555
362	610	576	501	622	658	341	517	715	436
307	465	458	301	474	478	583	434	573	837
468	430	207	371	582	846	514	562	569	714
453	564	581	624	539	427	372	609	316	427
435	662	537	589	795	683	747	469	455	709
766	527	688	639	614	717	405	780	858	328
593	513	624	715	536	508	277	502	427	816
650	595	701	491	207	541	609	430	630	558
492	550	552	550	726	583	367	403	410	627
387	395	675	602	606	476	253	534	466	448
513	528	456	726	520	599	769	528	492	499
719	541	654	368	625	344	636	452	429	405
615	547	292	590	383	505	585	325	519	624
494	530	231	404	633	719	477	454	508	515
540	363	409	565	542	489	273	509	543	669
403	707	305	589	734	576	553	466	332	632

Составить интервальный вариационный ряд. Записать эмпирическую функцию распределения и построить ее график. На одном чертеже изобразить гистограмму и полигон частот.

По сгруппированным данным вычислить выборочные числовые характеристики: среднее арифметическое, исправленную выборочную дисперсию, среднее квадратичное отклонение, коэффициент вариации, асимметрию, эксцесс, моду и медиану.

Заменив параметры генеральной совокупности соответственно их наилучшими выборочными числовыми характеристиками и используя χ^2-критерий Пирсона, на уровне значимости α=0,05 проверить две гипотезы о том, что изучаемая случайная величина ξ – величина транспортных затрат – распределена:

а) по нормальному закону распределения;

б) по равномерному закону распределения.

6. Распределение 60 предприятий по объему инвестиций в развитие производства ξ (млн.руб.) и получаемой за год прибыли η (млн.руб.) представлены в таблице:

η \ ξ	0–0,8	0,8–1,6	1,6–2,4	2,4–3,2	3,2–4,0	Итого
2–4	2	2				4
4–6	2	7	10			19
6–8		2	17	7		26
8–10			4	3	2	9
10–12					2	2
Итого	4	11	31	10	4	60

Необходимо:

1) Вычислить групповые средние $\bar{x}_i$ и $\bar{y}_j$, построить эмпирические линии регрессии;

2) Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

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

в) используя соответствующее уравнение регрессии, оценить полученную прибыль при объеме инвестиций 5 млн. руб.

ВАРИАНТ 8

(для студентов, номера личных дел которых оканчиваются цифрой 8)

1. Вероятность сбоя при получении денег в банкомате равна 0,001. Найти вероятность того, что из 5000 обращений число сбоев будет:

а) ровно 5;

б) не более 5.

2. Среднее число заявок, поступающих за сутки в диспетчерскую аварийной службы, равно 32. Оценить вероятность того, что в ближайшие сутки число поступивших заявок не превысит 40.

3.Полагая, что длина изготавливаемой детали есть нормально распределенная случайная величина с математическим ожиданием $M(X)$=10 см. и средним квадратическим отклонением $\sigma = 0,5$ см. Найти вероятность того, что длина наугад взятой детали заключена в интервале (8; 11). Какую отклонение длины от средней (по модулю) можно ожидать с вероятностью 0,95?

4. Туристическая фирма проводит опрос населения, чтобы выяснить, какое количество средств семья готова потратить на летний отдых. По схеме собственно случайной бесповторной выборки опрошено 200 семей. Результаты обследования приведены в таблице:

Предполагаемые затраты на отдых, т.р.	0менее 100	5100-150	1150-200	1200-250	2250-300	3300-350	бболее 350
Количество семей	888	332	118	112	116	114	220

Найти вероятность того, что средние данные по всему региону отличаются от средних данных в выборке не более, чем на 15 т. р. Сколько человек надо опросить, чтобы с той же вероятностью гарантировать ошибку в 10 т. р.? Население региона считать очень большим.

5. По схеме собственно-случайной бесповторной выборки проведено 10%-ное обследование предприятий одной из отраслей экономики в отчетном году с целью определения объема выпуска продукции (млн.руб.) Полученные данные представлены в таблице:

62,27	91,63	76,17	125,15	42,73	105,08	65,02	66,47	67,26	52,10
67,06	90,19	72,84	70,35	79,33	90,38	103,07	76,29	78,36	110,46
65,95	65,57	105,32	72,88	119,00	83,08	90,25	83,81	89,44	100,10
68,29	87,11	94,39	87,07	61,58	99,45	65,80	96,49	88,31	76,69

83,71	83,26	80,45	123,17	112,47	77,30	85,70	59,56	100,16	44,91
81,67	88,36	73,38	90,02	90,39	71,57	65,76	64,00	73,39	97,65
94,91	77,13	49,69	106,97	104,18	116,68	82,85	66,51	76,05	91,90
58,69	50,57	93,06	99,49	70,32	101,71	38,48	74,66	79,18	95,35
51,40	81,50	112,34	75,40	66,08	79,88	91,13	105,40	52,35	54,91
72,82	121,39	76,50	65,34	85,48	111,86	86,49	92,90	90,61	47,63
73,59	82,48	70,72	78,27	54,38	59,64	58,26	61,87	66,55	73,85
90,17	46,01	75,57	86,93	93,05	70,86	88,77	78,66	91,89	109,49
54,92	90,78	80,91	94,76	100,73	103,59	58,59	68,79	84,46	75,01
82,00	91,53	108,37	46,04	56,89	52,17	80,26	62,50	65,05	78,10
72,36	81,25	56,34	83,97	64,52	80,06	92,67	63,82	79,50	72,07
97,30	78,66	76,42	103,88	79,08	81,01	66,76	117,25	61,88	87,49

Составить интервальный вариационный ряд. Записать эмпирическую функцию распределения и построить ее график. На одном чертеже изобразить гистограмму и полигон частот.

По сгруппированным данным вычислить выборочные числовые характеристики: среднее арифметическое, исправленную выборочную дисперсию, среднее квадратичное отклонение, коэффициент вариации, асимметрию, эксцесс, моду и медиану.

Заменив параметры генеральной совокупности соответственно их наилучшими выборочными числовыми характеристиками и используя χ^2-критерий Пирсона, на уровне значимости α=0,05 проверить две гипотезы о том, что изучаемая случайная величина ξ – величина транспортных затрат – распределена:

а) по нормальному закону распределения;

б) по равномерному закону распределения.

6.Распределение 60 опрошенных студентов по количеству посещений бассейна в месяц ξ и ежемесячных внеплановых затрат η тыс. руб представлено в таблице.

ξ \ η	0–0,8	0,8–1,6	1,6–2,4	2,4–3,2	3,2–4,0	Итого
2–4	2	2				4
4–6	2	7	10			19
6–8		2	17	7		26
8–10			4	3	2	9
10–12					2	2
Итого	4	11	31	10	4	60

Необходимо:

1) Вычислить групповые средние $\overline{x}_i$ и $\overline{y}_j$, построить эмпирические линии регрессии;

2) Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

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

в) используя соответствующее уравнение регрессии, оценить среднее количество ежемесячных посещений бассейна при уровне внеплановых затрат, равном 2 тыс.руб.в месяц.

ВАРИАНТ 9

(для студентов, номера личных дел которых оканчиваются цифрой 9)

1. В городе 14% пенсионеров и среди них каждый двухсотый верит «некачественной» рекламе. Какова вероятность того, что хотя бы два пенсионера поверят рекламе, если в городе население 10000 человек?

2. Случайная величина X имеет нормальный закон распределения с параметрами a и σ^2. Найти параметры, если известно, что $P(X<1)=0{,}5$ и $P(-2<X<4)=0{,}9973$. Вычислить вероятность того, что значение случайной величины X окажется меньше 2.

3. Сумма вклада клиента сберегательного банка – это случайная величина, распределенная по биномиальному закону с математическим ожиданием 15 тыс. руб. и дисперсией 0,4.

1) Используя неравенство Чебышева, оценить вероятность того, что сумма вклада наудачу взятого вкладчика будет заключена в границах от 14 тыс. руб. до 16 тыс. руб.

2) Найти ту же вероятность, используя следствие из интегральной теоремы Муавра-Лапласа.

4. Для планирования бюджета предприятия на следующий год было проведено выборочное обследование использования амортизационного фонда. По схеме собственно- случайной бесповторной выборки среди 500 выплат из амортизационного фонда были отобраны 100 и получены следующие данные:

Величина выплаты (т.руб.)	Менее 100	100–200	200–300	300–400	400–500	500–600	Итого
Число выплат	3	13	33	26	17	8	100

Найти:

а) вероятность того, что средняя выплата отличается от средней выплаты в выборке не более чем на 15 т. руб.;

б) границы, в которых с вероятностью 0,9281 заключена доля выплат, величина которых не превосходит 400 руб.;

в) объем бесповторной выборки, при котором те же границы для доли (см. п. б)) можно гарантировать с вероятностью 0,9545.

5. С целью изучения роста производительности труда на предприятиях молочной промышленности по схеме собственно-случайной бесповторной выборки было обследовано 160 предприятий из 1500. Данные о величине роста производительности труда (%) представлены в таблице:

3,1	0,4	0,9	4,2	4,7	7,5	20	8,4	9,9	0,8
4,6	6,2	3,9	7,4	6,3	5,6	06	9,4	8,2	8,3
7,5	2,9	1,1	3,9	2,3	9,9	0,3	3,7	6,4	9,7
3,9	5,7	8,1	3,9	9,4	9,6	2,3	2,9	7,9	4,7
1,9	7,2	2,5	9,6	0,1	0,5	0,7	4,5	5,1	9,9
17	8,6	2,9	8,8	6,7	9,8	7,9	8,1	4,8	0,9
8,8	11	3,2	0,4	7,8	8,3	19	9,8	1,3	7,6
7,5	7,5	0,8	4,5	2,6	0,3	0,9	9,5	7,8	2,5
4,5	7,5	11	4,2	0,4	4,2	6,1	8,2	4,7	6,9
5,2	7,7	1,7	3,3	3,2	3,1	3,4	5,6	9,1	1,6
9,1	8,3	1,3	2,4	1,9	9,6	4,5	2,3	6,4	6,7

9,3	6,3	7,6	3,6	12	1,8	8,7	6,4	8,7	3,5
8,9	2,8	4,5	2,4	2,6	2,7	7,8	5,7	5,2	6,6
3,4	4,1	9,7	6,1	8,4	1,4	4,6	0,7	09	5,9
1,2	6,5	4,3	9,3	6,6	6,3	1,3	9,1	12	2,7
2,2	5,5	3,1	1,7	0,5	4,5	3,1	10	8,3	8,5

Составить интервальный вариационный ряд. Записать эмпирическую функцию распределения и построить ее график. На одном чертеже изобразить гистограмму и полигон частот.

По сгруппированным данным вычислить выборочные числовые характеристики: среднее арифметическое, исправленную выборочную дисперсию, среднее квадратичное отклонение, коэффициент вариации, асимметрию, эксцесс, моду и медиану.

Заменив параметры генеральной совокупности соответственно их наилучшими выборочными числовыми характеристиками и используя χ^2-критерий Пирсона, на уровне значимости α=0,05 проверить две гипотезы о том, что изучаемая случайная величина ξ – величина транспортных затрат – распределена:

а) по нормальному закону распределения;

б) по равномерному закону распределения.

6. Распределение 50 предприятий по численности работающих ξ (чел.) и объёму привлечённых инвестиций η (млн. руб.) представлено в таблице:

η \ ξ	1–2	2–3	3–4	4–5	5–6	Более 6	Итого
30–50	1	1	3				5
50–70		2	5	1			8
70-90		1	1	6	2	2	12
90–110			4	9			13
110-130			2	2	5		9
Более 130					2	1	3
Итого	1	4	15	18	9	3	50

Необходимо:

1) Вычислить групповые средние $\overline{x}_i$ и $\overline{y}_j$, построить эмпирические линии регрессии;

2) Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

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

в) используя соответствующее уравнение регрессии, оценить средний объём привлечённых инвестиций на предприятии с количеством работников, равным 100.

ВАРИАНТ 10

(для студентов, номера личных дел которых оканчиваются цифрой 0)

1. Человек, проходящий мимо киоска, покупает газету с вероятностью 0,2. Найти вероятность того, что среди 400 человек, прошедших мимо киоска в течение часа:

а) купят газету 90 человек;

б) не купят газету от 300 до 340 человек (включительно).

2. Плотность вероятности случайной величины X имеет вид:

$$\phi(x) = \begin{cases} 0 & при \quad x < 1, \\ \dfrac{1}{4} & при \quad 1 \le x \le b, \\ 0 & при \quad x > b. \end{cases}$$

Найти: а) параметр b; б) математическое ожидание и дисперсию случайной величины X; в) функцию распределения $F(x)$ и построить ее график.

Оценить с помощью неравенства Чебышева вероятность того, что случайная величина принимает значения на промежутке $[1{,}5;\,4{,}5]$. Вычислить эту вероятность с помощью функции распределения. Объяснить различие результатов.

3. Известно, что рост человека является нормально распределенной случайной величиной. В результате выборочного обследования средний рост мужчины оценен как 177 см, а дисперсия оказалась равной 50. Записать выражение плотности вероятности и функцию распределения случайной величины - роста мужчины. Найти вероятность того, что наудачу выбранный мужчина будет иметь рост: а) не менее 183 см, б) не более 180 см.

4. По схеме собственно-случайной бесповторной выборки проведено 10%-ное обследование строительных организаций региона по объему выполненных работ (млн. руб.). Результаты представлены в таблице:

Объем работ (млн. руб.)	Менее 56	56–60	60–64	64–68	68–72	Более 72	Итого
Число организаций	9	11	19	30	18	13	100

Найти:

а) границы, в которых с вероятностью 0,9973 заключен средний объем выполненных работ всех строительных организации региона;

б) вероятность того, что доля всех строительных организаций, объем работ которых не менее 60 млн. руб., отличается от доли таких организаций в выборке не более, чем на 0,05 (по абсолютной величине);

в) объем бесповторной выборки, при котором те же границы для среднего объема выполненных работ, (см. п. а)), можно гарантировать с вероятностью 0,9876.

5. По схеме собственно-случайной бесповторной выборки проведено 10%-ное обследование аптек региона по недельному объему продаж антибиотиков (тыс. руб.). Предполагая, что в регионе функционируют 1000 аптек, получены следующие данные:

748	449	713	602	775	661	147	676	108	488
612	641	761	660	642	794	636	924	859	866
839	573	510	597	735	135	435	759	645	695
597	795	671	596	922	694	556	572	668	776
729	656	738	941	702	707	479	610	783	698
824	877	572	887	649	984	668	857	616	498
682	716	749	706	667	865	896	697	519	841
838	838	711	609	740	433	714	940	848	561
609	837	715	766	451	603	639	673	613	821
784	665	534	751	580	748	753	629	686	724

Составить интервальный вариационный ряд. Записать эмпирическую функцию распределения и построить ее график. На одном чертеже изобразить гистограмму и полигон частот.

По сгруппированным данным вычислить выборочные числовые характеристики: среднее арифметическое, исправленную выборочную дисперсию, среднее квадратическое отклонение, коэффициент вариации, асимметрию, эксцесс, моду и медиану.

Заменив параметры генеральной совокупности соответственно их наилучшими выборочными числовыми характеристиками и используя χ^2-критерий Пирсона, на уровне значимости α=0,05 проверить две гипотезы о том, что изучаемая случайная величина ξ – недельный объем продаж антибиотиков – распределена:

а) по нормальному закону распределения;

б) по равномерному закону распределения.

6. Распределение 100 торговых павильонов по числу привлекаемых сезонных рабочих ξ (чел.) и их среднемесячной заработной плате на одного человека η (тыс. руб.) представлено в таблице:

η \ ξ	10–12	12–14	14–16	16–18	18–20	Более 20	Итого
1–3						10	10
3–5					6	15	21
5–7			10	11	8		29
7–9			8	3			11
9–11		5	6				11
11-13	5	9	4				18
Итого	5	14	28	14	14	25	100

Необходимо:

1) Вычислить групповые средние $\overline{x}_i$ и $\overline{y}_j$, построить эмпирические линии регрессии;

2) Предполагая, что между переменными ξ и η существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

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

в) используя соответствующее уравнение регрессии, оценить среднемесячную заработную плату одного сезонного рабочего в павильоне, привлекающем 10 сезонных рабочих.

§ 14. ЭКЗАМЕНАЦИОННЫЕ ВОПРОСЫ

Экзаменационные вопросы по дисциплине «Анализ данных» для бакалавров направлений 38.03.01 «Экономика» и 38.03.05 «Бизнес-информатика» составлены в соответствии с рабочей программой дисциплины.

Следует обратить внимание на то, что часть вопросов, например, вопрос о статистическом определении вероятности и теореме Бернулли, являются комплексными и направлены на проверку понимания связей между различными понятиями курса. Так статистическое определение вероятности, как правило, обсуждается на первой лекции, а теорема Бернулли доказывается на последней, а на экзамене студент должен понимать, что теорема Бернулли является теоретическим обоснованием справедливости и статистического, и классического определений вероятности в тех случаях, когда они оба применимы.

Часть вопросов помечена «*». Эти вопросы по усмотрению преподавателя могут быть включены в перечень обязательных и рассмотрены на лекции, могут быть предложены для самостоятельного изучения студентами, а могут быть представлены в форме реферативных докладов. Ниже приведён перечень вопросов:

1. Данные в экономике. Объекты, признаки и таблицы. Типы признаков в экономике и управлении: интервальные, порядковые, ранговые, дихотомические.
2. Инструменты описательной статистики. Измерение центра распределения. Измерение разброса данных.
3. Визуализация качественных признаков. Сводные таблицы и сводные диаграммы. Таблицы сопряженности и парадокс Симпсона. Иерархия признаков.
4. Предварительная обработка данных. Выбросы и их обработка. Пропущенные значения и их обработка. Повторяющиеся строки и их обработка. Синтетические признаки.

5. Определение вероятности. Случайные события, их виды. Операции над событиями как операции над множествами. Классическая вероятностная схема. Схема геометрических вероятностей. Статистическая вероятность. Аксиоматическое построение теории вероятностей. Теорема сложения вероятностей. Обобщенная теорема сложения вероятностей.
6. Условные вероятности. Условная вероятность. Независимость событий. Формула полной вероятности. Формула Байеса. Простейшие примеры применения теории вероятностей в экономике, управлении и финансах.
7. Последовательности испытаний. Биномиальная схема. Формула Бернулли. Формула Пуассона. Последовательности испытаний в экономике и управлении.
8. Определение случайной величины. Понятие случайной величины. Функция распределения случайной величины. Свойства функции распределения. Индикатор события как простейшая случайная величина. Функция распределения индикатора события.
9. Дискретные случайные величины и их важнейшие числовые характеристики. Дискретная случайная величина. Ряд распределения и функция распределения дискретной случайной величины. Математическое ожидание, дисперсия и среднее квадратическое отклонение дискретной случайной величины.
10. Биномиальный закон распределения.
11. Биномиальная модель ценообразования финансовых инструментов.
12. Геометрический закон распределения.
13. Закон распределения Пуассона.
14. Простейший поток событий.
15. Гипергеометрический закон распределения.
16. Сравнение случайных величин: отношение предпочтения, ожидаемая полезность, оптимальность по Парето. *

17. Абсолютно непрерывные случайные величины и их важнейшие числовые характеристики. Абсолютно непрерывная случайная величина. Функция распределения и функция плотности распределения абсолютно непрерывной случайной величины. Свойства функции плотности распределения. Математическое ожидание, дисперсия и среднее квадратическое отклонение абсолютно непрерывной случайной величины.
18. Равномерный закон распределения.
19. Показательный закон распределения.
20. Нормальный закон распределения.
21. Логарифмически нормальный закон распределения и ценообразование финансовых инструментов. *
22. Закон распределения Парето и задачи налогообложения. *
23. Законы распределения, важные в математической статистике (законы распределения Стьюдента, χ^2, Фишера — Снедекора). *
24. Смеси распределений. *
25. Начальные и центральные моменты случайной величины. Асимметрия и эксцесс случайной величины.
26. Квантили и процентные точки случайной величины.
27. Ценность под риском. *
28. Медиана и мода случайной величины.
29. Случайные векторы и условные законы распределения. Условный ряд распределения (для дискретных случайных величин), условная плотность распределения (для непрерывных случайных величин).
30. Условное математическое ожидание. Формула полного математического ожидания. Формула полной дисперсии.
31. Ковариация и коэффициент корреляции.
32. Портфель финансовых инструментов.*

33. Функции случайных величин. Функции одной случайной величины. Функции нескольких случайных величин. Формула композиции. Композиция равномерных случайных величин. *
34. Закон больших чисел. Массовые случайные явления в экономике. Теорема Чебышёва и оценка математического ожидания. Теорема Бернулли и оценка вероятности. Обсуждение условий статистической устойчивости.
35. Центральная предельная теорема. Теорема Леви.* Интегральная теорема Муавра — Лапласа. Место центральной предельной теоремы в изучении статистических закономерностей в экономике, финансах и управлении.
36. Математические основы теории страхования. *
37. Метод Монте-Карло. Моделирование случайных величин. *
38. Основы выборочного метода. Предмет и задачи математической статистики. Генеральная и выборочная совокупности. Случайная и конкретная выборки. Случайная повторная и случайная бесповторная выборка.
39. Соотношение между предельной ошибкой выборки, уровнем значимости (риском) и объемом выборки. Использование этого соотношения в организации выборочных обследований.
40. Оценка плотности распределения и функции распределения. Вариационный ряд. Выборочная случайная величина (статистический ряд распределения). Интервальный вариационный ряд. Полигон частот, кумулята. Оценка числовых характеристик генеральной случайной величины с помощью выборочной случайной величины. Выборочное среднее как оценка математического ожидания. Относительная частота как оценка вероятности. Выборочная дисперсия как оценка дисперсии.
41. Точечные оценки параметров. Понятие точечной оценки параметра генеральной совокупности. Свойства точечных оценок: состоятельность, несмещенность, эффективность.

42. Выборочное среднее как состоятельная, несмещенная и эффективная оценка математического ожидания генеральной случайной величины.
43. Смещенность выборочной дисперсии как оценки дисперсии генеральной случайной величины. Исправленная выборочная дисперсия как несмещенная и состоятельная оценка дисперсии генеральной случайной величины.
44. Методы построения точечных оценок: метод моментов, метод максимального правдоподобия. Примеры построения оценок параметров распределений случайных величин, применяемых в экономике и управлении.
45. Интервальные оценки параметров. Понятие интервальной оценки параметра генеральной совокупности. Точные интервальные оценки вероятности, математического ожидания, дисперсии и коэффициента корреляции. Поправка на конечный объем генеральной совокупности. Асимптотический подход к интервальному оцениванию.
46. Статистические гипотезы. Понятие статистической гипотезы. Виды статистических гипотез: параметрические и непараметрические, простые и сложные. Критерий проверки гипотезы, критическое множество. Проверка гипотез с помощью интервальных оценок. Ошибки первого и второго родов. Мощность критерия. Наиболее мощный критерий.
47. Проверка гипотезы о равенстве математического ожидания теоретическому значению. Проверка гипотезы о равенстве двух математических ожиданий.
48. Проверка гипотезы о равенстве дисперсии теоретическому значению. Проверка гипотезы о равенстве двух дисперсий.
49. Проверка гипотезы о равенстве вероятности события теоретическому значению. Проверка гипотезы о равенстве двух вероятностей.
50. Проверка гипотез о значимости коэффициента корреляции.

51. Критерии согласия. Критерий согласия χ^2 Пирсона. Критерий χ^2 Пирсона при неизвестных параметрах распределения.
52. Однофакторный дисперсионный анализ. *
53. Двухфакторный дисперсионный анализ. *
54. Таблицы сопряженности. Критерий χ^2 для проверки независимости компонент случайной величины. Критерий χ^2 для проверки однородности данных. *
55. Непараметрические критерии. Проверка гипотез на малых выборках. Критерий знаков. Распределение Вилкоксона и его критические границы. Непараметрическая точечная оценка математического ожидания. Непараметрическая интервальная оценка математического ожидания. Критерий Вилкоксона (парный критерий знаковых рангов). Примеры применения непараметрических критериев в экономике. *
56. Ранговая корреляция. Коэффициент ранговой корреляции Спирмена. Коэффициент ранговой корреляции Кендалла. Коэффициент конкордации. Проверка гипотез о значимости ранговых коэффициентов корреляции. Примеры использования ранговой корреляции в экономике. *
57. Задачи машинного обучения. Обучение с учителем и обучение без учителя. Классы задач машинного обучения: регрессия, классификация, кластерный анализ, поиск аномалий. Примеры задач машинного обучения в экономике, управлении и финансах. *
58. Линейная регрессия. Постановка задачи регрессионного анализа. Парная линейная регрессия. Множественная линейная регрессия. * Точечный и интервальный прогноз по модели регрессии. Примеры задач регрессии в экономике. Понятие о гетероскедастичности и автокорреляции. *

59. Классификация с обучением. Постановка задачи классификации с обучением. Логистическая регрессия. Понятие о деревьях решений. Кредитный скоринг. *
60. Кластерный анализ и поиск аномалий. Постановка задачи кластерного анализа. Метод К-средних. Сегментирование потребителей. Понятие о методах машинного обучения в задачах поиска аномалий. *

§ 15. ОБРАЗЦЫ ЭКЗАМЕНАЦИОННЫХ БИЛЕТОВ

Экзаменационный билет **ДЕМО 1**

1. Равномерный закон распределения, его определение, свойства и примеры. (15 баллов)
2. Спортсмен может взять высоту с вероятностью 0,4. Составить закон распределения случайной величины – числа попыток сделанных спротсменом до покорения высоты. Найти математическое ожидание и дисперсию. (15 баллов)
3. Вероятность сделать ошибку при передаче знака цифровой информации равна 0,0001. Найти вероятность того, что при передаче 6000 знаков будет: а) две ошибки; б) не более трех ошибок. (15 баллов)
4. Построить доверительный интервал, в котором с вероятностью 0,9545 заключена генеральная доля, если по результатам повторной выборки объема 100 получена выборочная доля $\omega = 0{,}3$. (15 баллов)

Экзаменационный билет **ДЕМО 2**

1. Нормальный (гауссовский) закон распределения. Геометрический и вероятностный смысл параметров нормального закона распределения. Стандартный нормальный закон распределения. Функция Гаусса, ее свойства и график. (15 баллов)
2. В коробку, где находилось 30 исправных батареек, по ошибке брошено 5 использованных. Из этой коробки наудачу достают 2 батарейки. Какова вероятность того, что: а) они обе исправны; б) хотя бы одна из них неисправна. (15 баллов)
3. Суточный расход воды в некотором населенном пункте – это случайная величина со средним квадратическим отклонением 10000 л. Оценить вероятность того, что в ближайшие сутки расход воды в этом населенном пункте отклонится от своего среднего значения не более чем на 25000л. (15 баллов)
4. Проектируется выборочное наблюдение, целью которого является установление среднего размера деталей в совокупности, состоящей из 10000 деталей. Требуемая точность 1 см. Произведенные пробные выборки дали наибольшую дисперсию, равную 49 см. Определить необходимую численность случайной бесповторной выборки, обеспечивающей заданную точность выборки, если ее надежность определяется вероятностью 0,95. (15 баллов)

ЛИТЕРАТУРА

Основная

1. Соловьёв В.И. *Анализ данных в экономике: теория вероятностей, прикладная статистика, обработка и визуализация данных в Microsoft Excel: учебник*. – Москва: КНОРУС, 2019. – 498 с.
2. Потемкин А.В., Фридман М.Н., Эйсымонт И.М. *Теория вероятностей и математическая статистика: учебное пособие.* М.: Финансовый университет, 2015.
3. Потемкин А.В., Эйсымонт И.М. *Анализ данных: учебное пособие*. – М.: Финансовый университет, 2014.
4. Кремер Н.Ш. *Теория вероятностей и математическая статистика*. М.: ЮНИТИ, 2003, 2004, 2007.
5. Геворкян П.С. *Теория вероятностей и математическая статистика: Курс лекций*/ П.С. Геворкян, А.В. Потемкин, И.М. Эйсымонт.— М.: Экономика, 2012.

Дополнительная

1. Браилов А.В., Солодовников А.С. *Сборник задач по курсу «Математика в экономике». Часть 3. Теория вероятностей*. М.:Финансы и статистика, 2010.
2. Денежкина И.Е., Орлова М.Г., Швецов Ю.Н. *Основы математической статистики. Учебно-методическое пособие для самостоятельной работы бакалавров.* М.: Финансовая академия при правительстве РФ, 2010.
3. Солодовников А.С., Бабайцев В.А., Браилов А.В., Шандра И.Г. *Математикав экономике. Учебник в 3 ч. Ч.3. Теория вероятностей и математическая статистика.* М.: Финансы и статистика, 2008.

Приложение 1

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

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3698
0,4	0,3683	0,3668	0,3652	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0395	0,0387	0,0379	0,0371	0,0363
2,2	0,0353	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001

Приложение 2-1

Таблица значений функции Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02892	0,02790	0,03188	0,03586
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,34134	0,34375	0,34614	0,34850	0,35083	0,35314	0,35543	035769	0,35993	0,36214
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41308	0,41466	0,41621	0,41774
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42786	0,42922	0,43056	0,43189
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49422	0,49506	0,49520
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49813	0,49819	0,49825	0,49831	0,49336	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995

Приложение 2-2

Таблица удвоенных значений функции Лапласа $\Phi(x)=\frac{2}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0,0797	0,0876	0,0955	0,1034	0,1113	0,1192	0,1271	0,1350	0,1428	0,1507
0,2	0,1585	0,1663	0,1741	0,1819	0,1897	0,1974	0,2051	0,2128	0,2205	0,2282
0,3	0,2358	0,2434	0,2510	0,2586	0,2661	0,2737	0,2812	0,2886	0,2960	0,3035
0,4	0,3108	0,3182	0,3255	0,3328	0,3401	0,3473	0,3545	0,3616	0,3688	0,3759
0,5	0,3829	0,3899	0,3969	0,4039	0,4108	0,4177	0,4245	0,4313	0,4381	0,4448
0,6	0,4515	0,4581	0,4647	0,4713	0,4778	0,4843	0,4907	0,4971	0,5035	0,5098
0,7	0,5161	0,5223	0,5285	0,5346	0,5407	0,5467	0,5527	0,5587	0,5646	0,5705
0,8	0,5763	0,5821	0,5878	0,5935	0,5991	0,6047	0,6102	0,6157	0,6211	0,6265
0,9	0,6319	0,6372	0,6424	0,6476	0,6528	0,6579	0,6629	0,6679	0,6729	0,6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	0,7287	0,7330	0,7373	0,7415	0,7457	0,7499	0,7540	0,7580	0,7620	0,7660
1,2	0,7699	0,7737	0,7775	0,7813	0,7850	0,7887	0,7923	0,7959	0,7984	0,8029
1,3	0,8064	0,8098	0,8132	0,8165	0,8198	0,8230	0,8262	0,8293	0,8324	0,8355
1,4	0,8385	0,8415	0,8444	0,8473	0,8501	0,8529	0,8557	0,8584	0,8611	0,8638
1,5	0,8664	0,8690	0,8715	0,8740	0,8764	0,8789	0,8812	0,8836	0,8859	0,8882
1,6	0,8904	0,8926	0,8948	0,8969	0,8990	0,9011	0,9031	0,9051	0,9070	0,9090
1,7	0,9109	0,9127	0,9146	0,9164	0,9181	0,9199	0,9216	0,9233	0,9249	0,9265
1,8	0,9281	0,9297	0,9312	0,9327	0,9342	0,9357	0,9371	0,9385	0,9392	0,9412
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9533
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	0,9643	0,9651	0,9660	0,9668	0,9676	0,9684	0,9692	0,9700	0,9707	0,9715
2,2	0,9722	0,9729	0,9736	0,9743	0,9749	0,9756	0,9762	0,9768	0,9774	0,9780
2,3	0,9786	0,9791	0,9797	0,9802	0,9807	0,9812	0,9817	0,9822	0,9827	0,9832
2,4	0,9836	0,9841	0,9845	0,9849	0,9853	0,9857	0,9861	0,9865	0,9869	0,9872
2,5	0,9876	0,9879	0,9883	0,9886	0,9889	0,9892	0,9895	0,9898	0,9901	0,9904
2,6	0,9907	0,9910	0,9912	0,9915	0,9917	0,9920	0,9922	0,9924	0,9926	0,9928
2,7	0,9931	0,9933	0,9935	0,9937	0,9939	0,9940	0,9942	0,9944	0,9946	0,9947
2,8	0,9949	0,9951	0,9952	0,9953	0,9955	0,9956	0,9958	0,9959	0,9960	0,9961
2,9	0,9963	0,9964	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	0,9981	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986
3,2	0,9986	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,3	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,4	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995	0,9995
3,5	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997	0,9997
3,6	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998	0,9998
3,7	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

Приложение 3

Таблица значений функции Пуассона $P_n(k) \approx \frac{\lambda^k}{k!} \cdot e^{-\lambda}$

k \ λ	**0,1**	**0,2**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,8**	**0,9**	**1,0**
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,0905	0,1637	0,2223	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659	0,3679
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1216	0,1438	0,1647	0,1839
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494	0,0613
4	0,0000	0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111	0,0153
5	0,0000	0,0000	0,0000	0,0001	0,0002	0,0003	0,0007	0,0012	0,0020	0,0031
6	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0003	0,0005

k \ λ	**2.0**	**3.0**	**4.0**	**5.0**	**6,0**	**7,0**	**8,0**	**9,0**	**10,0**
0	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0001
1	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011	0,0005
2	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050	0,0023
3	0,1805	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150	0,0076
4	0,0902	0,1681	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337	0,0189
5	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607	0,0378
6	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911	0,0631
7	0,0034	0,0216	0,0595	0,1045	0,1377	0,1490	0,1396	0,1171	0,0901
8	0,0009	0,0081	0,0298	0,0653	0,1033	0,1304	0,1396	0,1318	0,1126
9	0,0002	0,0027	0,0132	0,0363	0,0689	0,1014	0,1241	0,1318	0,1251
10	0,0000	0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186	0,1251
11	0,0000	0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970	0,1137
12	0,0000	0,0001	0,0006	0,0034	0,0113	0,0264	0,0481	0,0728	0,0948
13	0,0000	0,0000	0,0002	0,0013	0,0052	0,0142	0,0296	0,0504	0,0729
14	0,0000	0,0000	0,0001	0,0005	0,0022	0,0071	0,0169	0,0324	0,0521
15	0,0000	0,0000	0,0000	0,0002	0,0009	0,0033	0,0090	0,0194	0,0347
16	0,0000	0,0000	0,0000	0,0000	0,0003	0,0015	0,0045	0,0109	0,0217
17	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0021	0,0058	0,0128

Приложение 4

Критические точки распределения Стьюдента

Число степеней свободы *k*	Уровень значимости α (двусторонняя критическая область)					
	0,1	**0,05**	**0,02**	**0,01**	**0,002**	**0,001**
1	6,3138	12,7062	31,8205	63,6567	318,3088	636,6192
2	2,9200	4,3027	6,9646	9,9248	22,3271	31,5991
3	2,3534	3,1824	4,5407	5,8409	10,2145	12,9240
4	2,1318	2,7764	3,7469	4,6041	7,1732	8,6103
5	2,0150	2,5706	3,3649	4,0321	5,8934	6,8688
6	1,9432	2,4469	3,1427	3,7074	5,2076	5,9588
7	1,8946	2,3646	2,9980	3,4995	4,7853	5,4079
8	1,8595	2,3060	2,8965	3,3554	4,5008	5,0413
9	1,8331	2,2622	2,8214	3,2498	4,2968	4,7809
10	1,8125	2,2281	2,7638	3,1693	4,1437	4,5869
11	1,7959	2,2010	2,7181	3,1058	4,0247	4,4370
12	1,7823	2,1788	2,6810	3,0545	3,9296	4,3178
13	1,7709	2,1604	2,6503	3,0123	3,8520	4,2208
14	1,7613	2,1448	2,6245	2,9768	3,7874	4,1405
15	1,7531	2,1314	2,6025	2,9467	3,7328	4,0728
16	1,7459	2,1199	2,5835	2,9208	3,6862	4,0150
17	1,7396	2,1098	2,5669	2,8982	3,6458	3,9651
18	1,7341	2,1009	2,5524	2,8784	3,6105	3,9216
19	1,7291	2,0930	2,5395	2,8609	3,5794	3,8834
20	1,7247	2,0860	2,5280	2,8453	3,5518	3,8495
21	1,7207	2,0796	2,5176	2,8314	3,5272	3,8193
22	1,7171	2,0739	2,5083	2,8188	3,5050	3,7921
23	1,7139	2,0687	2,4999	2,8073	3,4850	3,7676
24	1,7109	2,0639	2,4922	2,7969	3,4668	3,7454
25	1,7081	2,0595	2,4851	2,7874	3,4502	3,7251
26	1,7056	2,0555	2,4786	2,7787	3,4350	3,7066
27	1,7033	2,0518	2,4727	2,7707	3,4210	3,6896
28	1,7011	2,0484	2,4671	2,7633	3,4082	3,6739
29	1,6991	2,0452	2,4620	2,7564	3,3962	3,6594
30	1,6973	2,0423	2,4573	2,7500	3,3852	3,6460
40	1,6839	2,0211	2,4233	2,7045	3,3069	3,5510
60	1,6706	2,0003	2,3901	2,6603	3,2317	3,4602
70	1,6669	1,9944	2,3808	2,6479	3,2108	3,4350
80	1,6641	1,9901	2,3739	2,6387	3,1953	3,4163
90	1,6620	1,9867	2,3685	2,6316	3,1833	3,4019
	0,0500	**0,0250**	**0,0100**	**0,0050**	**0,0010**	**0,0005**
	Уровень значимости α (односторонняя критическая область)					

Приложение 5-1

Критические точки распределения Фишера при уровне значимости α =0.01

(k_1— число степеней свободы большей дисперсии, k_2—число степеней свободы меньшей дисперсии)

k_2 \ k_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Приложение 5-2

Критические точки распределения Фишера при уровне значимости α =0.05

(k_1— число степеней свободы большей дисперсии, k_2—число степеней свободы меньшей дисперсии)

k_1 \ k_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,5	19,00	19,16	19,25	19:30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Приложение 6

Критические точки распределения χ^2

Число степеней свободы *k*	Уровень значимости α								
	0,01	**0,025**	**0,05**	**0,1**	**0,5**	**0,9**	**0,95**	**0,975**	**0,99**
1	6,634	5,023	3,841	2,705	0,454	0,015	0,003	0,0009	0,0002
2	9,210	7,377	5,991	4,605	1,386	0,210	0,102	0,0500	0,020
3	11,340	9,348	7,814	6,251	2,365	0,584	0,351	0,2150	0,114
4	13,270	11,14	9,487	7,779	3,356	1,063	0,710	0,4840	0,297
5	15,080	12,83	11,070	9,236	4,351	1,610	1,145	0,8310	0,554
6	16,810	14,44	12,590	10,640	5,348	2,204	1,635	1,2370	0,872
7	18,470	16,01	14,060	12,010	6,345	2,833	2,167	1,6890	1,239
8	20,090	17,53	15,500	13,360	7,344	3,489	2,732	2,1790	1,646
9	21,660	19,02	16,910	14,680	8,342	4,168	3,325	2,7000	2,087
10	23,200	20,48	18,300	15,980	9,341	4,865	3,940	3,2460	2,558
11	24,720	21,92	19,670	17,270	10,340	5,577	4,574	3,8150	3,053
12	26,210	23,33	21,020	18,540	11,340	6,303	5,226	4,4030	3,570
13	27,68	24,73	22,36	19,81	12,33	7,041	5,891	5,008	4,106
14	29,14	26,11	23,68	21,06	13,33	7,789	6,570	5,628	4,660
15	30,57	27,48	24,99	22,30	14,33	8,546	7,260	6,262	5,229
16	31,99	28,84	26,29	23,54	15,33	9,312	7,961	6,907	5,812
17	33,40	30,19	27,58	24,76	16,33	10,080	8,671	7,564	6,407
18	34,80	31,52	28,86	25,98	17,33	10,860	9,390	8,230	7,014
19	36,19	32,85	30,14	27,20	18,33	11,650	10,110	8,906	7,632
20	37,56	34,16	31,41	28,41	19,33	12,440	10,850	9,590	8,260
21	38,93	35,47	32,67	29,61	20,33	13,230	11,590	10,280	8,897
22	40,28	36,78	33,92	30,81	21,33	14,040	12,330	10,980	9,542
23	41,63	38,07	35,17	32,00	22,33	14,840	13,090	11,680	10,19
24	42,97	39,36	36,41	33,19	23,33	15,650	13,840	12,400	10,85
25	44,31	40,64	37,65	34,38	24,33	16,470	14,610	13,110	11,52
26	45,64	41,92	38,88	35,56	25,33	17,290	15,370	13,840	12,19
27	46,96	43,19	40,11	36,74	26,33	18,110	16,150	14,570	12,87
28	48,27	44,46	41,33	37,91	27,33	18,930	16,920	15,300	13,56
29	49,58	45,72	42,55	39,08	28,33	19,760	17,700	16,040	14,25
30	50,89	46,97	43,77	40,25	29,33	20,590	18,490	16,790	14,95