

ПРИЛОЖЕНИЕ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ  
ИМЕНИ В.Ф. УТКИНА»

Кафедра «Электронные вычислительные машины»

## **ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ**

### **Б1.В.ДВ.01.01 «Анализ данных»**

Направление подготовки

02.03.03 Математическое обеспечение и администрирование информационных систем

Профиль

«Математическое обеспечение и администрирование информационных систем»

Уровень подготовки

Бакалавриат

Квалификация выпускника – бакалавр

Форма обучения – очная, очно-заочная

Рязань 2024

## 1. ОБЩИЕ ПОЛОЖЕНИЯ

Оценочные материалы – это совокупность учебно-методических материалов (контрольных заданий, описаний форм и процедур проверки), предназначенных для оценки качества освоения обучающимися данной дисциплины как части ОПОП.

Цель – оценить соответствие знаний, умений и владений, приобретенных обучающимся в процессе изучения дисциплины, целям и требованиям ОПОП в ходе проведения промежуточной аттестации.

Основная задача – обеспечить оценку уровня сформированности компетенций, закрепленных за дисциплиной.

Промежуточный контроль по дисциплине осуществляется путем проведения экзамена. Форма проведения экзамена – билеты с письменным ответом на два теоретических вопроса и одним практическим заданием. При необходимости, проводится устная беседа с обучаемым для уточнения оценки.

Выполнение заданий на практических занятиях в течение семестра и заданий на самостоятельную работу является обязательным условием для допуска к экзамену.

## 2 ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ

Сформированность компетенции (или ее части) в рамках освоения данной дисциплины оценивается по трехуровневой шкале:

- 1) пороговый уровень является обязательным для всех обучающихся по завершении освоения дисциплины;
- 2) продвинутый уровень характеризуется превышением минимальных характеристик сформированности компетенций по завершении освоения дисциплины;
- 3) эталонный уровень характеризуется максимально возможной выраженностью компетенций и является важным качественным ориентиром для самосовершенствования.

### Уровень освоения компетенций, формируемых дисциплиной:

**Описание критериев и шкалы оценивания теоретических вопросов и практического задания:**

| Шкала оценивания                | Критерий  |
|---------------------------------|---|
| 5 баллов<br>(эталонный уровень) | выставляется студенту, который дал полные ответы на вопросы, показал глубокие систематизированные знания, смог привести примеры, решил практическую задачу, ответил на дополнительные вопросы преподавателя |
| 4 балла (продвинутый уровень)   | выставляется студенту, который дал преимущественно полные ответы на вопросы, решил практическую задачу, но на некоторые дополнительные вопросы преподавателя ответил только с помощью наводящих вопросов    |
| 3 балла (пороговый уровень)     | выставляется студенту, который дал неполные ответы на вопросы в билете, показал в основном верный ход решения задачи и смог ответить на дополнительные вопросы только с помощью преподавателя               |
| 2 балла                         | выставляется студенту, который не смог ответить на  |

### 3 ПАСПОРТ ОЦЕНОЧНЫХ МАТЕРИАЛОВ ПО ДИСЦИПЛИНЕ

| Контролируемые разделы (темы) дисциплины (результаты по разделам)                                 | Код контролируемой компетенции (или её части) | Наименование оценочного средства |
|---|---|----------------------------------|
| Тема 1. Введение. Типы данных. Методы сбора и анализа данных.                                     | ПК-6  | Экзамен                          |
| Тема 2. Выборочный метод анализа данных. Характеристики выборки.                                  | ПК-6  | Экзамен                          |
| Тема 3. Параметры распределения и их оценка. Точечные и интервальные оценки.                      | ПК-6  | Экзамен                          |
| Тема 4. Проверка статистических гипотез.  | ПК-6  | Экзамен                          |
| Тема 5. Корреляционный анализ.  | ПК-6  | Экзамен                          |
| Тема 6. Двумерная дискретная случайная величина.  | ПК-6  | Экзамен                          |
| Тема 7. Хранилище данных. Data Mining. Многомерная модель данных. Интеллектуальный анализ данных. | ПК-6  | Экзамен                          |

### 4. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ

#### 4.1. Промежуточная аттестация в форме экзамена

| Коды компетенций | Результаты освоения ОПОП<br>Содержание компетенций  |
|------------------|---|
| ПК-6             | Способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования |

#### а) типовые практические задания на экзамен:

##### Задание 1

В таблице приведена реализация выборки из генеральной совокупности зарегистрированной средней скорости передачи данных по каналу связи в Мб/с. Выполнить группировку. Для определения количества интервалов и их границ применить формулу Стреджесса. Подсчитать количество наблюдений в каждой из групп, определить их частоту.

|    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|
| 22 | 27 | 28 | 18 | 29 | 29 | 16 | 12 | 35 |
| 24 | 13 | 23 | 34 | 18 | 17 | 29 | 15 | 32 |
| 23 | 18 | 21 | 20 | 22 | 21 | 17 | 11 | 16 |
| 16 | 23 | 28 | 21 | 24 | 23 | 25 | 14 | 19 |
| 15 | 19 | 15 | 36 | 16 | 13 | 18 | 20 | 23 |
| 20 | 19 | 11 | 17 | 22 | 27 | 34 | 20 | 15 |
| 13 | 28 | 27 | 22 | 18 | 11 | 21 | 13 | 21 |
| 18 | 17 | 31 | 21 | 25 | 22 | 16 | 26 | 33 |

|    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|
| 27 | 24 | 22 | 17 | 12 | 24 | 24 | 26 | 19 |
| 15 | 15 | 10 | 17 | 17 | 15 | 26 | 27 | 28 |
| 31 | 23 | 24 | 22 | 24 | 28 | 21 | 27 | 26 |
| 15 | 13 | 16 | 19 | 28 | 29 | 28 | 32 | 30 |
| 13 | 12 | 26 | 19 | 27 | 31 | 27 | 24 | 33 |
| 16 | 14 | 18 | 32 | 19 | 18 | 20 | 24 | 24 |
| 18 | 17 | 21 | 19 | 15 | 22 | 21 | 18 | 13 |
| 21 | 12 | 29 | 18 | 30 | 17 | 13 | 23 | 29 |
| 20 | 11 | 25 | 19 | 18 | 29 | 18 | 24 | 21 |
| 27 | 22 | 25 | 17 | 15 | 13 | 20 | 24 | 17 |
| 33 | 27 | 23 | 24 | 26 | 30 | 24 | 29 | 23 |
| 16 | 21 | 30 | 27 | 18 | 14 | 19 | 15 | 12 |
| 28 | 23 | 25 | 18 | 20 | 27 | 15 | 34 | 27 |
| 18 | 29 | 28 | 14 | 20 | 15 | 21 | 21 | 28 |
| 18 | 16 | 24 | 25 | 28 | 25 | 23 | 19 | 14 |
| 14 | 21 | 21 | 16 | 28 | 19 | 15 | 27 | 24 |
| 15 | 16 | 19 | 12 | 27 | 22 | 35 | 25 | 17 |
| 22 | 21 | 24 | 22 | 21 | 18 | 19 | 21 | 27 |
| 32 | 14 | 21 | 22 | 25 | 31 | 32 | 24 | 15 |
| 25 | 24 | 16 | 24 | 18 | 25 | 30 | 30 | 19 |
| 24 | 29 | 16 | 26 | 18 | 20 | 27 | 25 | 14 |
| 22 | 24 | 14 | 15 | 30 | 19 | 32 | 17 | 28 |
| 16 | 27 | 17 | 23 | 27 | 26 |    |    |    |

**Ответ в виде таблицы (может быть выполнен отдельными расчетами):**

| Величина       | Размах $R$    | Число интервалов $m$ | Длина интервала $h$ |            |
|----------------|---------------|----------------------|---------------------|------------|
| Значение       | 26            | 9                    | 2,9                 |            |
| Интервалы      | Левая граница | Правая граница       | $n_i$               | $\omega_i$ |
| [8,55; 11,45)  | 8,55          | 11,45                | 5                   | 0,018      |
| [11,45; 14,35) | 11,45         | 14,35                | 24                  | 0,087      |
| [14,35; 17,25) | 14,35         | 17,25                | 45                  | 0,163      |
| [17,25; 20,15) | 17,25         | 20,15                | 46                  | 0,167      |
| [20,15; 23,05) | 20,15         | 23,05                | 46                  | 0,167      |
| [23,05; 25,95) | 23,05         | 25,95                | 34                  | 0,123      |
| [25,95; 28,85) | 25,95         | 28,85                | 40                  | 0,145      |
| [28,85; 31,75) | 28,85         | 31,75                | 21                  | 0,076      |
| [31,75; 34,65) | 31,75         | 34,65                | 12                  | 0,043      |
| [34,65; 36,1)  | 34,65         | 36,1                 | 3                   | 0,011      |

**Задание 2**

По данным предварительной группировки выборки, а также определения интервальных частот и частостей найти и построить на диаграммах эмпирическую плотность, эмпирическую функцию распределения и полигон.

| Величина       | Размах $R$    | Число интервалов $m$ | Длина интервала $h$ |            |
|----------------|---------------|----------------------|---------------------|------------|
| Значение       | 26            | 9                    | 2,9                 |            |
| Интервалы      | Левая граница | Правая граница       | $n_i$               | $\omega_i$ |
| [8,55; 11,45)  | 8,55          | 11,45                | 5                   | 0,018      |
| [11,45; 14,35) | 11,45         | 14,35                | 24                  | 0,087      |
| [14,35; 17,25) | 14,35         | 17,25                | 45                  | 0,163      |
| [17,25; 20,15) | 17,25         | 20,15                | 46                  | 0,167      |
| [20,15; 23,05) | 20,15         | 23,05                | 46                  | 0,167      |
| [23,05; 25,95) | 23,05         | 25,95                | 34                  | 0,123      |
| [25,95; 28,85) | 25,95         | 28,85                | 40                  | 0,145      |
| [28,85; 31,75) | 28,85         | 31,75                | 21                  | 0,076      |
| [31,75; 34,65) | 31,75         | 34,65                | 12                  | 0,043      |
| [34,65; 36,1)  | 34,65         | 36,1                 | 3                   | 0,011      |

**Ответы:** По данным таблицы построим гистограммы частот и частостей (рисунки 1 и 2 соответственно).

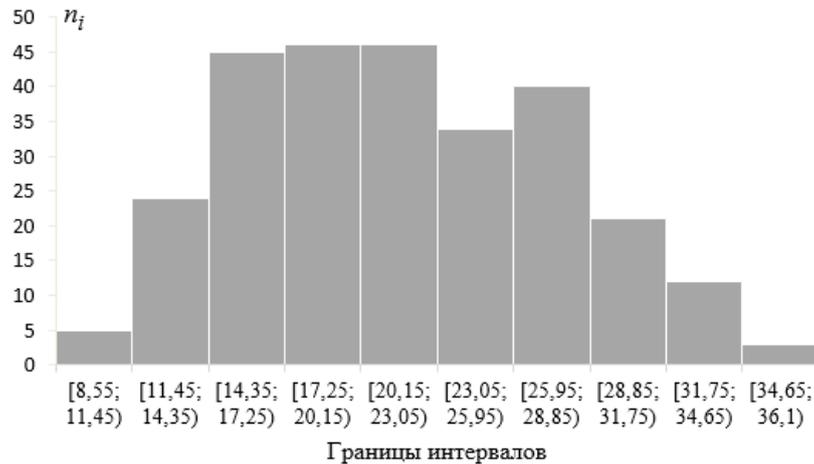


Рисунок 1 – Гистограмма частот

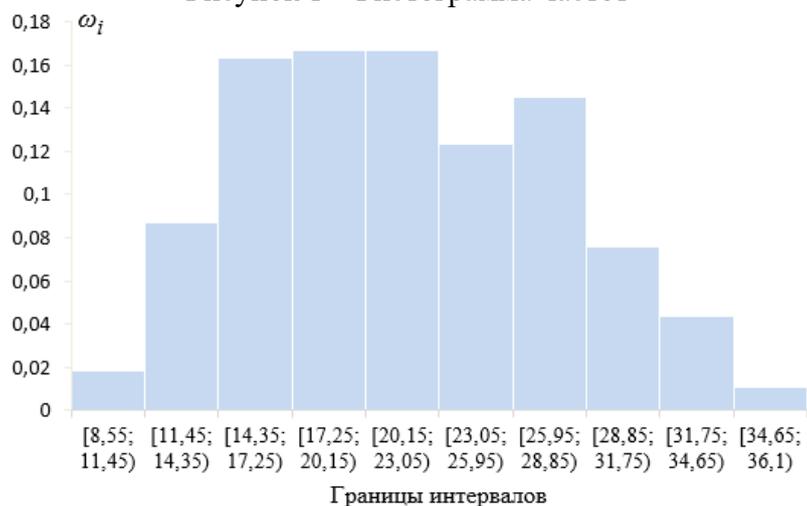


Рисунок 2 – Гистограмма частостей

По данным таблицы в одной системе координат построим гистограмму и полигон относительных частот – частостей (рисунок 3).

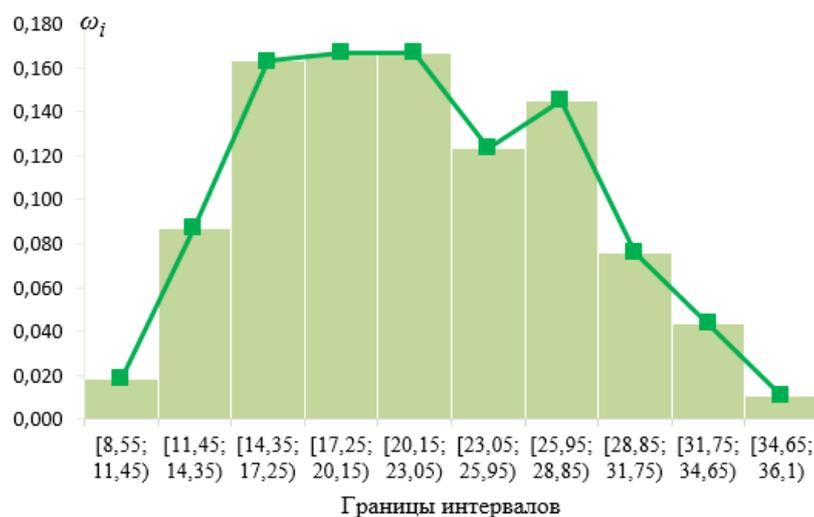


Рисунок 3 – Гистограмма и полигон частотей

Функция распределения  $F^*(x)$  является «ступенчатой», т.е. у нее имеются разрывы в точках, которым соответствуют правые границы интервалов. Величина скачка равна относительной частоте интервала.

$$F^*(x) = \begin{cases} 0; & x \leq x_1, \\ \omega_1; & x_1 < x \leq x_2, \\ \omega_1 + \omega_2; & x_2 < x \leq x_3, \\ \vdots \\ \sum_{i=1}^{m-1} \omega_i; & x_i < x \leq x_{i+1}, \\ 1; & x > x_m. \end{cases}$$

Найдем значения эмпирической функции распределения и представим их в табличном виде. Затем построим график эмпирической функции распределения на основе данных, для чего предварительно запишем аналитически данную функцию.

| Интервалы     | $\omega_i$ | $F^*(x)$ |
|---------------|------------|----------|
| $< 8,55$      | 0,000      | 0,000    |
| 8,55 – 11,45  | 0,018      | 0,018    |
| 11,45 – 14,35 | 0,087      | 0,105    |
| 14,35 – 17,25 | 0,163      | 0,268    |
| 17,25 – 20,15 | 0,167      | 0,435    |
| 20,15 – 23,05 | 0,167      | 0,601    |
| 23,05 – 25,95 | 0,123      | 0,725    |
| 25,95 – 28,85 | 0,145      | 0,870    |
| 28,85 – 31,75 | 0,076      | 0,946    |
| 31,75 – 34,65 | 0,043      | 0,989    |
| 34,65 <       | 0,011      | 1,000    |

$$F^*(x) = \begin{cases} 0; & x \leq 8,55, \\ 0,018; & 8,55 < x \leq 11,45, \\ 0,105; & 11,45 < x \leq 14,35, \\ 0,268; & 14,35 < x \leq 17,25, \\ 0,435; & 17,25 < x \leq 20,15, \\ 0,602; & 20,15 < x \leq 23,05, \\ 0,725; & 23,05 < x \leq 25,95, \\ 0,87; & 25,95 < x \leq 28,85, \\ 0,946; & 28,85 < x \leq 31,75, \\ 0,989; & 31,75 < x \leq 34,65, \\ 1; & x > 34,65. \end{cases}$$

Графики эмпирической функции распределения и кумуляты представлены на рисунке 4.

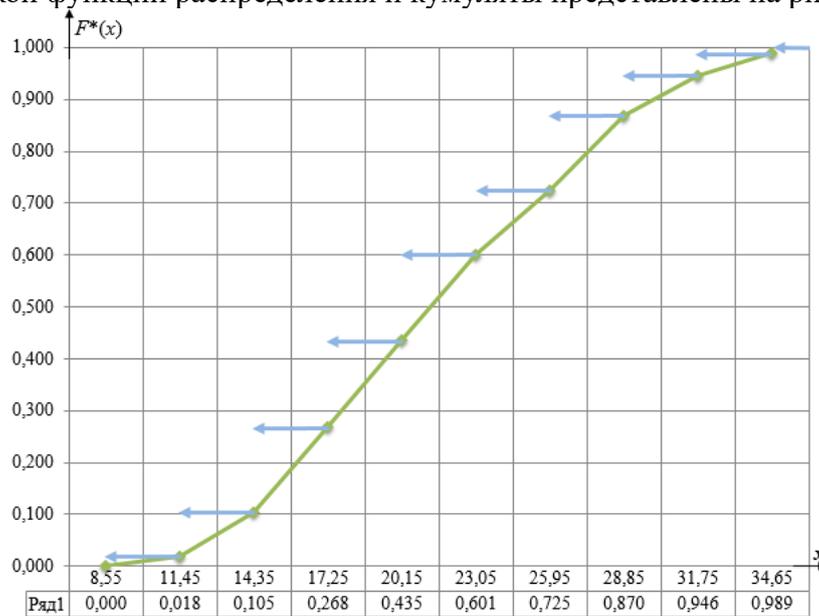


Рисунок 4 – Эмпирическая функция распределения и кумулята

### Задание 3

По данным предварительной группировки выборки, а также определения интервальных частот и частостей найти точечные оценки математического ожидания и дисперсии.

| Величина       | Размах $R$    | Число интервалов $m$ | Длина интервала $h$ |            |
|----------------|---------------|----------------------|---------------------|------------|
| Значение       | 26            | 9                    | 2,9                 |            |
| Интервалы      | Левая граница | Правая граница       | $n_i$               | $\omega_i$ |
| [8,55; 11,45)  | 8,55          | 11,45                | 5                   | 0,018      |
| [11,45; 14,35) | 11,45         | 14,35                | 24                  | 0,087      |
| [14,35; 17,25) | 14,35         | 17,25                | 45                  | 0,163      |
| [17,25; 20,15) | 17,25         | 20,15                | 46                  | 0,167      |
| [20,15; 23,05) | 20,15         | 23,05                | 46                  | 0,167      |
| [23,05; 25,95) | 23,05         | 25,95                | 34                  | 0,123      |
| [25,95; 28,85) | 25,95         | 28,85                | 40                  | 0,145      |
| [28,85; 31,75) | 28,85         | 31,75                | 21                  | 0,076      |

|                |       |       |    |       |
|----------------|-------|-------|----|-------|
| [31,75; 34,65) | 31,75 | 34,65 | 12 | 0,043 |
| [34,65; 36,1)  | 34,65 | 36,1  | 3  | 0,011 |

**Ответы:**

Точечная оценка *математического ожидания*  $m$  генеральной совокупности – выборочное среднее значение  $\bar{x}$  :

$$v_1 = \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^m x_i \cdot n_i = \sum_{i=1}^m x_i \cdot \omega_i,$$

где  $n$  – объем выборки,  $x_i$  – середины частичных интервалов  $[z_i; z_i+h)$ , а  $n_i$  – частоты ( $\omega_i$  – частости) попадания элементов выборки в частичные интервалы.

Точечная оценка *выборочной дисперсии*  $D$  генеральной совокупности – выборочное среднее квадратов отклонений значений выборки от выборочного среднего  $\bar{x}$  :

$$\mu_2 = D = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 \cdot n_i}{n} = \sum_{i=1}^m (x_i - \bar{x})^2 \cdot \omega_i.$$

также можно представить ее в виде  $\mu_2 = D = \frac{1}{n} \cdot \sum_{i=1}^m x_i^2 \cdot n_i - \bar{x}^2 = \sum_{i=1}^m x_i^2 \cdot \omega_i - \bar{x}^2 = \overline{x^2} - \bar{x}^2$ .

Точечная оценка *выборочного среднего квадратического отклонения* – величина, численно равная  $\sigma = \sqrt{D}$ .

**Расчеты:**

| Границы интервалов | $n_i$ | $\omega_i$ | Средины интервалов $x_i$ | $x_i \cdot n_i$ | $(x_i - \bar{x})^2 \cdot n_i$ |
|--------------------|-------|------------|--------------------------|-----------------|-------------------------------|
| [8,55; 11,45)      | 5     | 0,018      | 10                       | 50              | 686,58                        |
| [11,45; 14,35)     | 24    | 0,087      | 12,9                     | 309,6           | 1866,26                       |
| [14,35; 17,25)     | 45    | 0,163      | 15,8                     | 711             | 1576,13                       |
| [17,25; 20,15)     | 46    | 0,167      | 18,7                     | 860,2           | 419,04                        |
| [20,15; 23,05)     | 46    | 0,167      | 21,6                     | 993,6           | 0,64                          |
| [23,05; 25,95)     | 34    | 0,123      | 24,5                     | 833             | 263,11                        |
| [25,95; 28,85)     | 40    | 0,145      | 27,4                     | 1096            | 1291,31                       |
| [28,85; 31,75)     | 21    | 0,076      | 30,3                     | 636,3           | 1546,59                       |
| [31,75; 34,65)     | 12    | 0,043      | 33,2                     | 398,4           | 1581,98                       |
| [34,65; 36,1)      | 3     | 0,011      | 35,375                   | 106,12          | 559,52                        |
| Суммы              | 276   | 1          | –                        | 5994,22         | 9791,17                       |

1. Математическое ожидание  $\bar{x} = \frac{5994,22}{276} \approx 21,72$ .
2. Дисперсия  $D = \frac{9791,17}{276} \approx 35,48$ .
3. Среднее квадратическое отклонение  $\sigma = \sqrt{35,48} \approx 5,96$ .

**Задание 4**

В таблице приведена выборка из генеральной совокупности нормально распределенной случайной величины суточного увеличения объема данных на сервере (Мбайт) с *известным СКО*, равным 3,1 Мбайт. Для коэффициента доверия  $\gamma=0,98$  найти интервальную оценку математического ожидания случайной величины (суточного увеличения объема данных на сервере) (Мбайт).

|    |    |    |    |    |    |
|----|----|----|----|----|----|
| 44 | 44 | 43 | 52 | 47 | 40 |
| 41 | 46 | 45 | 48 | 46 | 43 |
| 47 | 49 | 41 | 41 | 50 | 48 |
| 47 | 48 | 49 | 51 | 51 | 51 |

|    |    |    |    |    |    |
|----|----|----|----|----|----|
| 46 | 43 | 42 | 48 | 48 | 50 |
|----|----|----|----|----|----|

**Ответы:**

Объем выборки  $n=30$ , а  $\sum_{i=1}^n x_i = 1389$ , точечная оценка математического ожидания

$$\bar{x} = \frac{1389}{30} = 46,3.$$

Интервальная оценка математического ожидания:  $m \in \left( \bar{x} - \frac{z_{\gamma/2} \cdot \sigma}{\sqrt{n}}; \bar{x} + \frac{z_{\gamma/2} \cdot \sigma}{\sqrt{n}} \right)$ , где  $\sigma=3,1$  (Мбайт),  $z_{\gamma/2}=2,32$  – значение функции Лапласа уровня  $\gamma/2=0,49$  (**Разрешается использовать табличные значения – справочные значения**).

Тогда значение выражения  $\frac{z_{\gamma/2} \cdot \sigma}{\sqrt{n}} = \frac{2,32 \cdot 3,1}{\sqrt{30}} \approx 1,31$ . С учетом этого  $m \in (46,3 - 1,31; 46,3 + 1,31) = (44,99; 47,61)$ .

**Задание 5**

В таблице приведена выборка из генеральной совокупности нормально распределенной случайной величины – количество одновременно нагруженных ЭВМ в вычислительном кластере с **неизвестным СКО**. Для заданного коэффициента доверия  $\gamma=0,95$  найти интервальную оценку математического ожидания случайной величины (среднего количества одновременно нагруженных ЭВМ).

|    |    |    |    |    |    |
|----|----|----|----|----|----|
| 44 | 33 | 41 | 38 | 35 | 43 |
| 36 | 36 | 37 | 40 | 38 | 38 |
| 36 | 42 | 41 | 44 | 39 | 37 |
| 39 | 45 | 43 | 41 | 42 | 45 |
| 45 | 40 | 38 | 36 | 44 | 40 |
| 39 | 36 | 36 | 39 | 38 |    |
| 35 | 35 | 39 | 42 | 37 |    |

**Ответы:**

Объем выборки  $n=40$ , а  $\sum_{i=1}^n x_i = 1572$ , точечная оценка математического ожидания

$$\bar{x} = \frac{1572}{40} = 39,3.$$

Интервальная оценка математического ожидания:  $m \in \left( \bar{x} - \frac{t(\gamma, n-1) \cdot S}{\sqrt{n}}; \bar{x} + \frac{t(\gamma, n-1) \cdot S}{\sqrt{n}} \right)$ ,  $t(\gamma; n-1)=2,02$  – квантиль распределения Стьюдента для уровня значимости  $\alpha=0,05$  и числа степеней свободы  $k=n-1=39$  (**Разрешается использовать табличные значения – справочные значения**).

Найдем  $(x_i - \bar{x})^2$  в соответствии с данными выборки:

| $(x_i - \bar{x})^2$ |
|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| 22,09               | 39,69               | 2,89                | 1,69                | 18,49               | 13,69               |
| 10,89               | 10,89               | 5,29                | 0,49                | 1,69                | 1,69                |
| 10,89               | 7,29                | 2,89                | 22,09               | 0,09                | 5,29                |
| 0,09                | 32,49               | 13,69               | 2,89                | 7,29                | 32,49               |
| 32,49               | 0,49                | 1,69                | 10,89               | 22,09               | 0,49                |
| 0,09                | 10,89               | 10,89               | 0,09                | 1,69                |                     |
| 18,49               | 18,49               | 0,09                | 7,29                | 5,29                |                     |

Исправленная выборочная дисперсия  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{408,40}{39} \approx 10,47$ , а  $S = \sqrt{S^2} = \sqrt{10,47} \approx 3,24$ .

Доверительный интервал математического ожидания  $m \in \left( 39,3 - \frac{2,02 \cdot 3,24}{\sqrt{40}}; 39,3 + \frac{2,02 \cdot 3,24}{\sqrt{40}} \right) = (38,26; 40,33)$ .

### Задание 6

Для оценки стабильности интернет-соединения каждый час передается 2000 пакетов данных, при этом из подвергнутых контролю 500 пакетов 130 пакетов было принято с потерями. Требуется определить точечную и интервальную оценку вероятности передачи пакета с потерями с заданным коэффициентом доверия  $\gamma=0,9$  и какое наименьшее количество пакетов необходимо передать, чтобы относительная частота пакетов, принятых с потерями, отличалась от вероятности потери не более, чем на 0,05 (5 %).

**Ответы:** Точечная оценка вероятности передачи пакета с потерями  $\omega = \frac{m}{n} = \frac{130}{500} = 0,26$ .

Интервальная оценка вероятности передачи пакета с потерями:  $(\omega - \Delta; \omega + \Delta)$ .

Для различных способов отбора:

- 1) повторный -  $\Delta = t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n}}$ ,
- 2) бесповторный -  $\Delta = t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n} \cdot \left(1 - \frac{n}{N}\right)}$ ,

где  $t = z_{\gamma/2}$  – аргумент функции Лапласа уровня  $\gamma/2 = 0,9/2 = 0,45$ . По справочным таблицам определяем, что  $t = z_{\gamma/2} = (1,64 + 1,65)/2 = 1,645$ .

Для повторного способа организации предельная ошибка  $\Delta = z_{\gamma/2} \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n}} = 1,645 \cdot \sqrt{\frac{0,26 \cdot (1-0,26)}{500}} \approx 0,0323$ , а интервальная оценка вероятности передачи пакета с потерями  $(0,26 - 0,0323; 0,26 + 0,0323) = (0,2277; 0,2923)$ .

Для бесповторного способа организации предельная ошибка  $\Delta = 1,645 \cdot \sqrt{\frac{0,26 \cdot (1-0,26)}{500} \cdot \left(1 - \frac{500}{2000}\right)} \approx 0,017$ , а интервальная оценка вероятности передачи пакета с потерями  $(0,26 - 0,017; 0,26 + 0,017) = (0,243; 0,277)$ .

Наименьшее количество пакетов, которое необходимо передать, чтобы относительная частота пакетов, принятых с потерями, отличалась от вероятности потери не более, чем на 0,05 (5 %).

Имеем  $\Delta = 0,05$  (5 %),  $z_{\gamma/2} = 1,645$ ,  $\omega = 0,26$ .

Для повторного способа организации выборки получим  $n = \frac{z_{\gamma/2}^2 \cdot \omega \cdot (1-\omega)}{\Delta^2} = \frac{1,645^2 \cdot 0,26 \cdot (1-0,26)}{0,05^2} \approx 209$  пакетов, а для бесповторного способа организации

$n = \frac{z_{\gamma/2}^2 \cdot \omega \cdot (1-\omega) \cdot N}{N \cdot \Delta^2 + z_{\gamma/2}^2 \cdot \omega \cdot (1-\omega)} = \frac{1,645^2 \cdot 0,26 \cdot (1-0,26) \cdot 2000}{2000 \cdot 0,05^2 + 1,645^2 \cdot 0,26 \cdot (1-0,26)} \approx 189$  пакетов.

### Задание 7

Из генеральной совокупности измерений зависимости скорости беспроводного соединения (Мбит/с) от дальности между передатчиком и приемником данных произведена выборка объемом 400 измерений, по которой получен выборочный коэффициент корреляции 0,78. Найти с надежностью  $\gamma=0,95$  доверительный интервал для коэффициента корреляции генеральной совокупности.

**Ответы:**

Показатель достоверности коэффициента корреляции  $t_r = \frac{r}{\Delta_r} = \frac{r}{\frac{1-r^2}{\sqrt{n}}} = \frac{r \cdot \sqrt{n}}{1-r^2}$ , где  $r=0,78$ ,

а  $n=400$ . Тогда  $t_r = \frac{0,78 \cdot \sqrt{400}}{1-0,78^2} \approx 40$ . Поскольку  $t_r > 3$ , то значение коэффициента корреляции можно считать достоверным.

Объем выборки  $n=400 > 100$ , доверительный интервал коэффициента корреляции генеральной совокупности принадлежит интервалу  $\rho \in \left( r - t_\gamma \cdot \frac{1-r^2}{\sqrt{n}}; r + t_\gamma \cdot \frac{1-r^2}{\sqrt{n}} \right)$ , где  $t_\gamma = 1,96$  – значение статистики распределения Стьюдента (**Разрешается использовать табличные значения – справочные значения**) с числом степеней свободы  $k=n-2=400-2=398$  при уровне значимости  $\alpha=1-\gamma=1-0,95=0,05$ . С учетом этого доверительный интервал  $\rho \in \left( 0,78 - 1,96 \cdot \frac{1-0,78^2}{\sqrt{400}}; 0,78 + 1,96 \cdot \frac{1-0,78^2}{\sqrt{400}} \right) \approx (0,742; 0,818)$ .

Выборочный коэффициент корреляции можно считать статистически значимым, так как  $t_r \approx 40 > 1,96 = t_\gamma$ .

### Вопросы к экзамену по дисциплине «Анализ данных»

1. Выборочный метод в статистических исследованиях (выборочная совокупность, размах, частичные интервалы, шаг разбиения, формула Стерджесса, частота, частость).
2. Выборочный метод в статистических исследованиях (гистограмма частот и частостей, полигон частот, эмпирическая функция распределения, кумулята, пример).
3. Выборочный метод в статистических исследованиях (моменты случайной величины: математическое ожидание, дисперсия, асимметрия, эксцесс, выборочный коэффициент вариации, ошибки асимметрии и эксцесса, общая форма записи моментов случайных величин).
4. Интервальные оценки числовых характеристик выборки (оценка математического ожидания с известным  $\sigma$ ).
5. Интервальные оценки числовых характеристик выборки (оценка математического ожидания с неизвестным  $\sigma$ ).
6. Интервальные оценки числовых характеристик выборки (оценка  $\sigma$ ).
7. Способы формирования выборок и отбора элементов в совокупность. Способы расчета объема выборки.
8. Способы формирования выборок и отбора элементов в совокупность. Средняя ошибка выборки.
9. Интервальная оценка выборочного коэффициента корреляции.
10. Проверка статистических гипотез.
11. Проверка гипотезы о равенстве выборочной средней генеральному среднему значению  $m$  нормально распределенной совокупности  $N(m; \sigma)$ .
12. Проверка гипотезы о равенстве числовых характеристик генеральных совокупностей. Проверка гипотезы о равенстве заданному числу дисперсии нормально распределенной случайной величины (одновыборочный  $\chi^2$  – критерий).
13. Проверка гипотезы о равенстве числовых характеристик генеральных совокупностей. Сравнение дисперсий двух выборок, имеющих нормальное распределение.
14. Проверка гипотезы о равенстве числовых характеристик генеральных совокупностей. Сравнение средних значений двух выборок, имеющих нормальное распределение.

15. Проверка гипотезы о равенстве числовых характеристик генеральных совокупностей. Доверительный интервал для разности средних.
16. Проверка гипотезы о равенстве долей признака. Сравнение генеральной доли со стандартом (нормативом).
17. Проверка гипотезы о равенстве долей признака. Проверка гипотезы о равенстве долей признака в двух совокупностях.
18. Оценка значимости коэффициента (индекса) корреляции.
19. Проверка гипотезы о виде распределения. Критерий согласия Пирсона.
20. Корреляционный и регрессионный анализ.
21. Задачи корреляционного анализа.
22. Вычисление коэффициента корреляции. Сила корреляционной связи. Парный коэффициент линейной корреляции.
23. Вычисление коэффициента корреляции. Сила корреляционной связи. Корреляционное отношение.
24. Вычисление коэффициента корреляции. Сила корреляционной связи. Коэффициент ранговой корреляции Спирмена.
25. Виды парных регрессионных зависимостей. Линейная регрессия.
26. Виды парных регрессионных зависимостей. Гиперболическая регрессия.
27. Виды парных регрессионных зависимостей. Экспоненциальная (показательная) регрессия.
28. Виды парных регрессионных зависимостей. Квадратичная регрессия.
29. Проверка адекватности модели и оценка значимости ее параметров.
30. Проверка статистической значимости коэффициентов линейной регрессии.
31. Доверительные интервалы для коэффициентов линейной регрессии.
32. Множественная регрессия.