Практическая работа №1. Первичная обработка данных

2023-02-05

# **Цель работы**

Данная практическая работа ставит перед собой задачу познакомить с процессом вычисления стандартных описательных статистик для выборок данных раазличных типов и научить использовать методы агрегации статистических данных с целью получения новых знаний об их эмпирическом распределении.

# **Постановка задачи**

Для выполнения задачи 1 раздела необходимо разбиться на две подгруппы.

Студенты первой подгруппы должны собрать со всей учебной группы данные о росте в сантиметрах и данные о месяцах рождения.

Студенты второй подгруппы должны собрать со всей учебной группы данные о росте в сантиметрах и данные о загаданном случайном целом числе на интервале [0; 8].

Для целочисленных данных необходимо:

1. Построить вариационный ряд распределения абсолютных и относительных частот появления событий по выборке дискретных данных.
2. Построить полигон относительных частот для событий вариационного ряда.
3. Вычислить эмпирическую функцию распределения и построить её график.
4. Рассчитать выборочные описательные статистики:

* среднее ;
* математическое ожидание ;
* дисперсию ;
* стандартное отклонение ;
* среднеквадратическое отклонение ;
* медиану зафиксированной выборки ;
* первый и третий квартиль ;
* межквартильный размах ;
* коэффициент вариации .

Для вещественных данных необходимо:

1. Рассчитать число групп , необходимых для квантования исходных данных по правилу Стёрджесса.
2. Вычислить значения границ групп для значений выборки по правилу фиксированной величины интервала.
3. Построить вариационный ряд для выборки интервальных данных.
4. Построить гистограмму распределения относительных частот для рассчитанных интервалов выборки.
5. Вычислить эмпирическую функцию распределения и построить её график.
6. Рассчитать выборочные описательные статистики:

* среднее ;
* математическое ожидание ;
* дисперсию c использованием выборочного среднего ;
* стандартное отклонение ;
* среднеквадратическое отклонение ;
* медиану зафиксированной выборки ;
* первый и третий квартиль ;
* межквартильный размах ;
* коэффициент вариации .

# **Пример расчета**

На рассмотрение выносится набор данных или выборка с двумя переменными, целочисленного и вещественного типа:

Выборочное среднее целочисленных данных рассчитывается по соотношению:

Математическое ожидание целочисленных данных рассчитывается по определению математического ожидания дискретного ряда:

Дисперсия целочисленных данных рассчитывается по определению дисперсии дискретного ряда с использованием выборочного среднего:

Стандартное отклонение с использованием дисперсии:

Среднеквадратическое отклонение:

Медиана выборки серединное значение отсортированной выборки:

# **Вопросы на защиту практической работы**