**Отчёт об исследовании “Возраста детей из аудитории компании 7 Футов Тур”**

**Первичная обработка данных**

Собранные данные

|  |
| --- |
| **Возраст детей** |
| 2 |
| 3 |
| 4 |
| 4 |
| 4 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 6 |
| 6 |
| 6 |
| 6 |
| 6 |
| 6 |
| 6 |
| 6 |
| 6 |
| 6 |
| 7 |
| 7 |
| 7 |
| 7 |
| 7 |
| 7 |
| 7 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 8 |
| 9 |
| 9 |
| 9 |
| 9 |
| 9 |
| 9 |
| 9 |
| 10 |
| 10 |
| 10 |
| 10 |
| 10 |
| 10 |
| 10 |
| 10 |
| 10 |
| 11 |
| 11 |
| 11 |
| 11 |
| 11 |
| 11 |
| 11 |
| 11 |
| 11 |
| 11 |
| 11 |
| 11 |
| 11 |
| 11 |
| 11 |
| 12 |
| 12 |
| 12 |
| 12 |
| 12 |
| 12 |
| 13 |
| 14 |
| 14 |
| 14 |
| 14 |
| 14 |
| 14 |
| 14 |
| 15 |
| 19 |
| 20 |
| 21 |
| 23 |
| 27 |

Гистограмма по данным

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **3. Рассчитайте следующие значения:** |  |  |  |  |
| 3.1 Выборочное среднее |  |  |  |  |
| 3.2 Выборочное стандартное отклонение |  |  |  |  |
| **Характеристика** | **Обозначение** | **Значение** | **Что характерезует** | **Как вычисляется** |
| Выборочное среднее | x¯ | 9.6 | ср. ар. всех эл. набора | ср. ар. всех эл. набора |
| Среднеквадратичное (стандартное) отклонение ДЛЯ ВЫБРОРОЧНОЙ ДИСПЕРСИИ | s | 4.1 | Характеризует разброс значений относительно среднего (Математического ожидания). Она используется, т.к. единицы её измерения совпадают с единицами измерения данных набора. | √(s²) |

**4. Данные дискретны или непрерывны? Обоснуйте.**

|  |
| --- |
| **Мои данные о возрасте - дискретны.** |
| Т.к. они представлены из множества целых чисел. И продолжительность жизни человека ограничена. |

**5. Словами опишите как строится гистограмма.**

|  |  |
| --- | --- |
| Шаг | **5. Алгоритм построения гистограммы** |
| 1 | Собрать данные |
| 2 | Выявить максимальное и минимальное значения данных. |
| 3 | Определить диапазон (размах) гистограммы вычитая из максимального значения данных минимальное |
|  |  |
| 4 | Полученный диапазон разделить на интервалы, предварительно определив их число (обычно 7-15 в зависимости от числа показателей) |
| 5 | Все данные распределить по интервалам в порядке возрастания: левая граница первого интервала должна быть меньше наименьшего из имеющихся значений |
| 6 | Подсчитать частоту каждого интервала (Сколько данных в него входит) |
| 7 | По полученным данным построить гистограмму - столбчатую диаграмму, высота столбиков которой соответствует частоте попадания данных в каждый из интервалов: |

**6. Есть ли потенциальные выбросы?**

Да, есть.

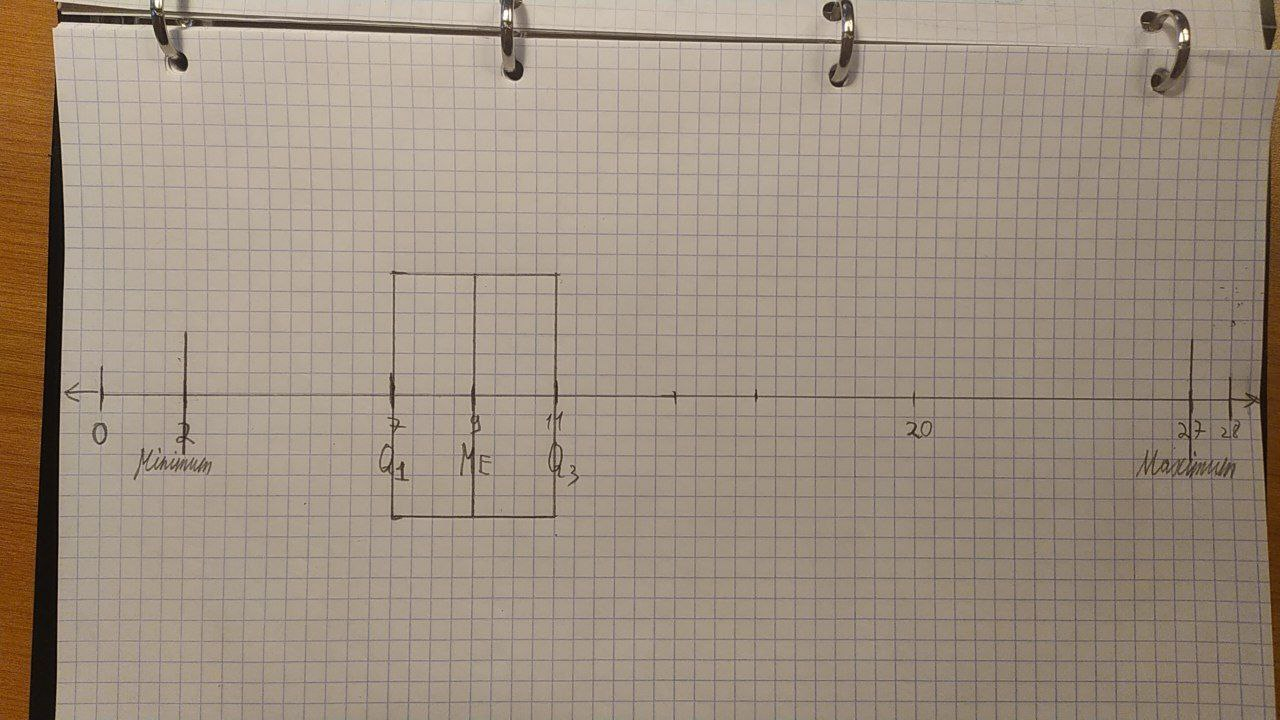
В таблице собранных данных умеренные выбросы выделены зеленым, а экстремальные – крсным цветом.

Подробные расчёты в отчёте Excel.

**Исследовательский анализ данных**

|  |  |
| --- | --- |
| **1. Определите следующие значения** |  |
| **Характеристика** | **Значение** |
| Min | 2 |
| Медиана | 9 |
| Max | 27 |
| Q₁ | 7 |
| Q₃ | 11 |
| IQR | 4 |
| 90-й процентиль | 14 |

**2. Постройте коробковую диаграмму**



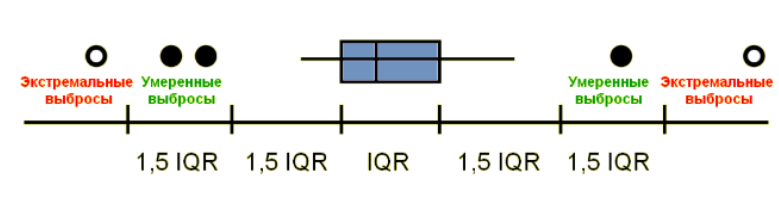
**3. Что форма диаграммы говорит о концентрации данных?**

Коробчатая диаграмма показывает, что наибольшая часть детей находится в возрасте от 7 до 11 лет (между 1 и 3 квартилями)

**4. Как с помощью коробковой диаграммы определить, есть ли потенциальные выбросы?**

Необходимо отметить на ней места, которые ограничивают диапазоны выбросов.

Если имеются данные, которые лежат в этих диапазонах, то они и являются выбросами



**5. Как стандартное отклонение помогает определить концентрацию данных и наличие потенциальных выбросов?**

Стандартное отклонение s определяет, насколько далеко может быть значение данных от среднего значения набора данных.

Если s маленькое, то данные сильно сконцентрированны.

Если s велико, то вероятность наличия выбросов сильно возрастает.

**6. Что представляет собой IQR в этой задаче?**

IQR = 4

"В моей задача IQR это разность между первыми 25 % детей и последними 25% детей по возрасту.

Т.е. он говорит, что разница между первыми 25% и 75% детей по возрасту составляет 4 года."

7. Найдите значение, которое составляет 1.5 стандартных отклонения:

Среднее - медиана - x¯ = 9.6

Стандартное отклонение s = 4.1

а. выше среднего \_\_\_\_\_\_\_ 15.7

b. ниже среднего \_\_\_\_\_\_\_ 3.5