

Описательная статистика

Построение гистограммы с заданными границами интервалов

```
In data.hist(bins=[value1, value2, value3, value4, ..., valueN])
```

Импорт библиотеки высокоуровневых математических функций

```
In import numpy as np
```

Нахождение дисперсии

```
In import numpy as np  
np.var(x)
```

Извлечение квадратного корня

```
In import numpy as np  
np.sqrt(x)
```

Словарь

Количественная (численная) переменная принимает числовое значение в диапазоне

Категориальная (качественная) переменная принимает значение из ограниченного набора

Непрерывная переменная количественная переменная, которая может принимать любое численное значение

Дискретная переменная количественная переменная, которая может принимать строго определённые значения

Плотность частоты величина, равная высоте столбца гистограммы, площадь которого отображает частоту непрерывной переменной

Плотностная гистограмма гистограмма, использующая в качестве переменной-столбца плотность частоты

Метрики локации данных характерные значения выборки, по значениям которых можно судить, где примерно расположен набор данных на числовой оси

Алгебраическая метрика локации среднее значение

Структурная метрика локации медиана

Дисперсия среднее значение квадрата отклонения случайной величины от её среднего значения. Вычисляется по формуле:

$$\sigma^2 = \frac{\sum (\mu - x_i)^2}{n}$$

Стандартное отклонение квадратный корень из дисперсии. Вычисляется по формуле:

$$\sigma = \sqrt{\frac{\sum (\mu - x_i)^2}{n}}$$

Правило трёх сигм практически все значения (около 99%) находятся в промежутке:

$$(\mu - 3\sigma, \mu + 3\sigma)$$

Скошенность асимметричность датасета

Датасет с положительной скошенностью (скошенностью вправо) датасет, гистограмма значений которого асимметрична, а для её симметрии необходимо добавить значения справа

Датасет с отрицательной скошенностью (скошенностью влево) датасет, гистограмма значений которого асимметрична, а для её симметрии необходимо добавить значения слева