

Базовые понятия
математической статистики
(2)

Описательная статистика

Локализация

- Среднее значение
- Медиана
- Мода

Дисперсия

- Перцентиль
- Межквартильный размах
- Размах признака
- Дисперсия
- Стандартное отклонение
- Коэффициент вариации

Среднее значение

- Mean (англ.)
- Обозначение – M
- Формула = $\sum x/n$

Медиана (1)

- Median (англ.)
- Обозначение: Me
- Медиана делит ряд на ДВЕ равные части
- Как найти: проранжировать значения от минимального к максимальному и выбрать стединое

Медиана (1)

- Медиана для чётного ряда – выбрать срединное значение из ранжированного ряда
- Медиана для нечетного ряда – суммировать два срединных значения и разделить на два

Пример: 1,5,2,8,7
Ранжир.: 1,2,5,7,8
 $Me=5$

Пример: 1,5,10,2,8,7
Ранжир.: 1,2,5,7,8,10
 $Me=(5+7)/2=6$

Мода (1)

- Mode (англ.)
- Обозначение Mo
- Мода - это самое часто встречающееся значение в шкале

Мода (2)

- Найдите наиболее частое значение в представленной шкале

Класс интенсивности работы	Количество данных
2	12
3	48
4	56
5	60
6	14
Всего	190

Перцентиль (1)

- Перцентиль указывает как данные распределены от минимального значения к максимальному
- Формула расчета $i \% = (i/100) \times (n+1)$

Перцентиль (2)

- Например, в ряду 2,4,5,6,7,9,10

25 перцентиль: $i=25$, $n=7$

$$(25/100) \times (n+1) = (25/100) \times (7+1) = 2$$

$$25\% = 4$$

- В ряду 2,4,5,6,7,9

25 перцентиль: $i=25$, $n=6$

$$(25/100) \times (n+1) = (25/100) \times (6+1) = 1,76$$

25% = 4 , т.к место под номером 1,76 между значениями «2» и «4» и ближе к «4»

Перцентиль (3)

- 25 –й перцентиль – *первый квартиль*
- 50 –й перцентиль – *второй квартиль*
- 75 –й перцентиль – *третий квартиль*

• В ряду : 2,4,5,6,7,9,10

25 –й перцентиль – 4

50 –й перцентиль – 6

75 –й перцентиль – 9

Размах признака

- Разница между наблюдением с минимальным значением и максимальным
- Пример: Вес
- Размах= $100-60=40$

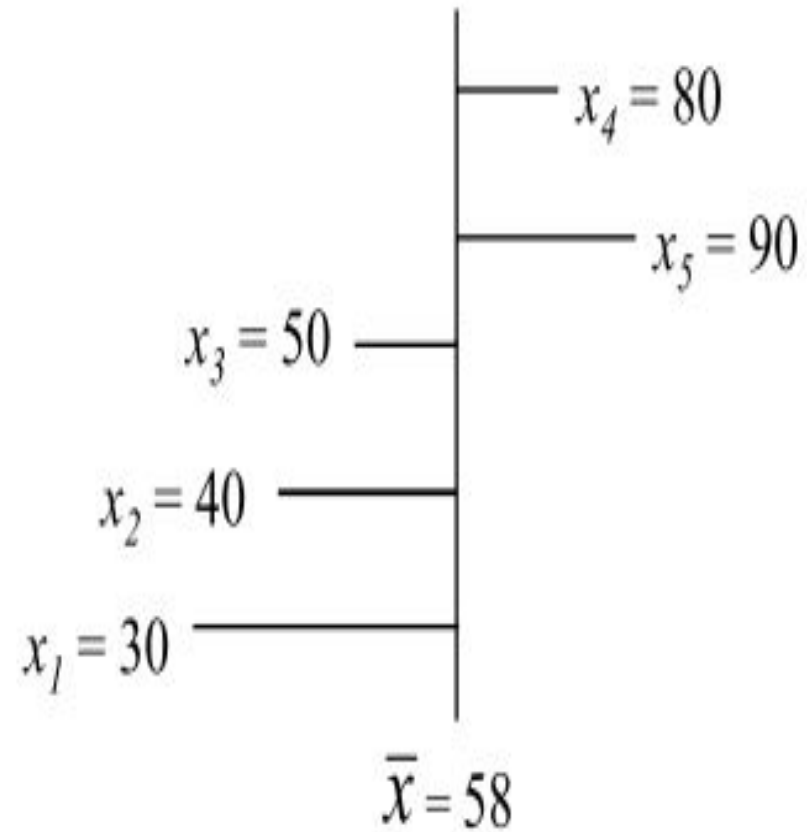
№	Вес
1	75
2	60
3	80
4	72
5	85
6	100
7	90
8	85

Стандартное отклонение (1)

- Измерение того как среднее значение представляет данные
- Малое стандартное отклонение указывает на то, что данные близки к среднему
- Большое стандартное отклонение указывает на то, что данные далеки от среднего значения

Стандартное отклонение (2)

- Отклонение пульса в пяти измерениях



Стандартное отклонение (3)

Шаг	Формула
1. Определяем среднее значение	$\bar{x} = \sum x/n$
2. Вычисляем насколько каждое наблюдение отстоит от среднего	$(x_i - \bar{x})$
3. Вычисляем квадрат разницы	$(x_i - \bar{x})^2$
4. Вычисляем сумму квадратов разниц	$\sum (x_i - \bar{x})^2$
5. Находим среднее значение квадратов разниц - ДИСПЕРСИЮ	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$
6. Вычисляем квадратный корень из дисперсии	$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

Стандартное отклонение (4)

x	$(x - \bar{x})$	$(x - \bar{x})^2$
170	-10	100
180	0	0
190	10	100
540 ($\bar{x} = 180$)	0	200

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{200}{3 - 1}} = \underline{10}$$

$n=3$

Стандартное отклонение (5)

x	$(x - \bar{x})$	$(x - \bar{x})^2$
179	-1	1
180	0	0
181	1	1
540 ($\bar{x} = 180$)	0	2

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{2}{3 - 1}} = 1$$

$n=3$

Степени свободы

- Количество наблюдений варьирующихся свободно
- Внутри всех значений n свободно варьируются только $(n-1)$, оставшееся значение неизменно
- $(n-1)$ значения изменяются т.о. именно им присвоена степень свободы

Коэффициент вариации

- характеристика стандартного отклонения

$$Cv = \frac{s}{X} \times 100 \%$$

- < 10% малый к.в.
- 10-20% средний к.в.
- >20% высокий к.в.