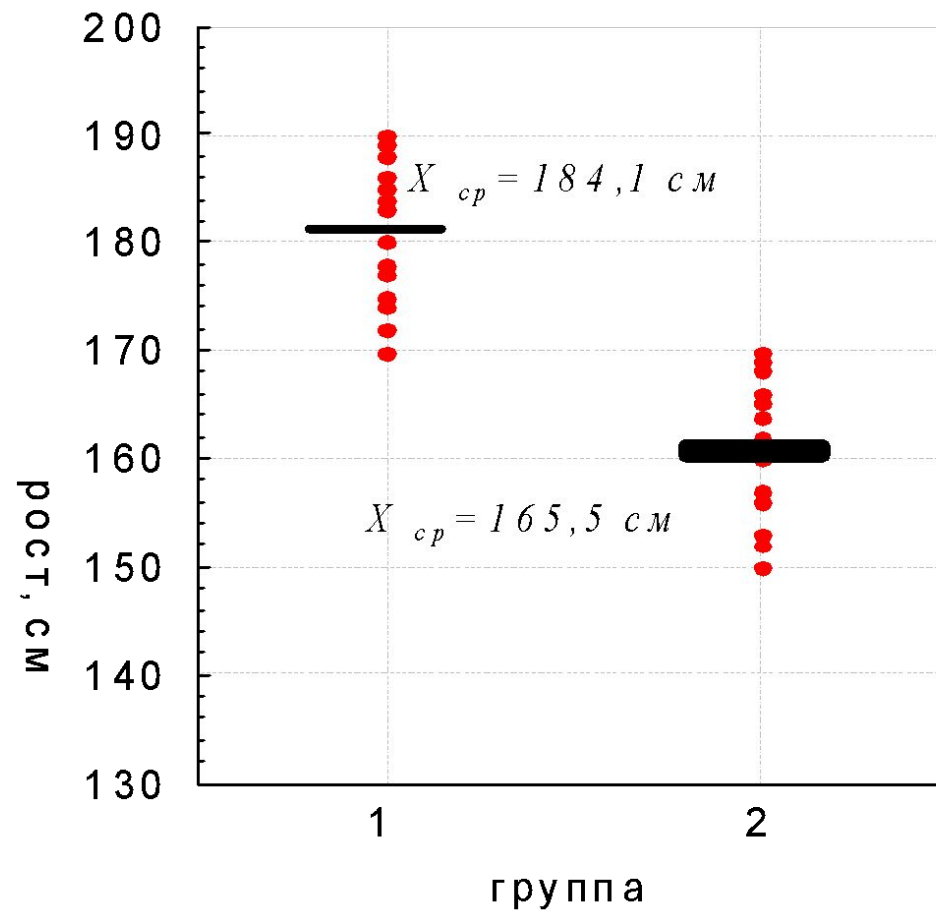


Вычисление статистических характеристик случайных величин

Для описания случайных величин используется целый ряд числовых статистических характеристик.

Среднее значение: \bar{X} показывает, где на числовой оси располагается изучаемая совокупность, другими словами это некоторое значение случайной величины, возле которого группируются все другие.

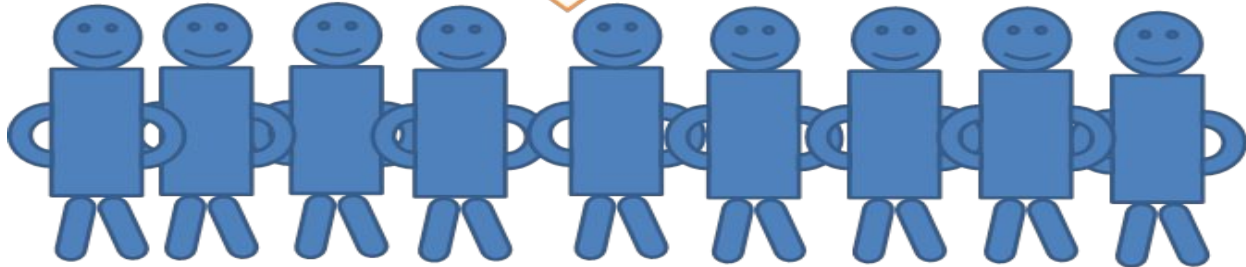


Медиана (Me) – это значение случайной величины, которое делит ранжированную выборку на две равные части. Если число объектов выборки четное, то медиана равна среднему двух соседних центральных значений. Половина объектов выборки имеет значение меньше медианы, половина – больше медианы.

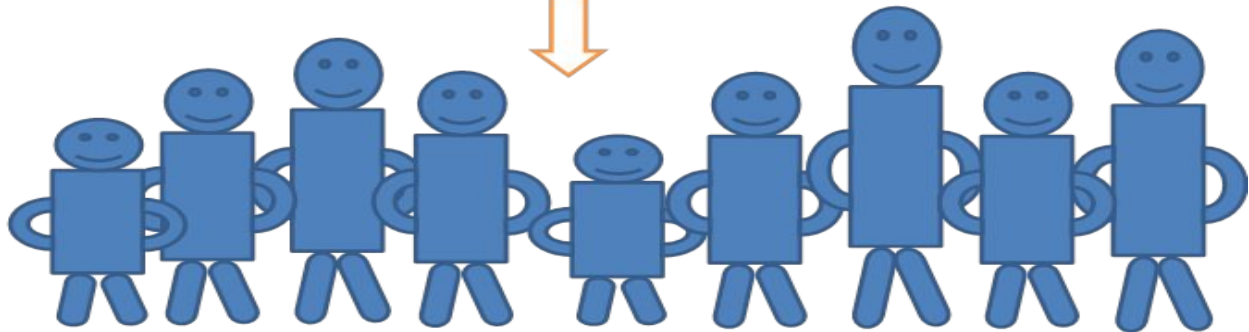
Мода (Mo) – наиболее часто встречающееся значение случайной величины. Для того, чтобы определить моду все значения выборки выстраиваются в ранжированный ряд (по возрастанию или по убыванию). Может быть несколько значений моды, может ее и не быть.

п	рост мальчиков	рост девочек
1	167	160
2	168	161
3	170	163
4	170	164
5	170	165
6	172	165
7	176	165
8	178	168
9	182	168
10	186	169
11	188	170

$$\bar{X} = 178 \text{ cm}$$



$$\bar{X} = 178 \text{ cm}$$



Дисперсия (D) – характеристика разброса значений случайной величины относительно среднего значения

Стандартное отклонение (среднеквадратичное отклонение) – также является характеристикой разброса, введена для того, чтобы избавиться от квадрата единицы измерения

Коэффициент вариации представляет собой относительную меру разброса, выраженную в процентах $V\%$ (иногда обозначается C_v)

$$V\% = \frac{s}{\bar{X}} \times 100\%$$

Коэффициент вариации:

используют для сравнения разброса двух и более признаков, имеющих различные единицы измерения.

он позволяет судить об однородности совокупности:

< 17% – абсолютно однородная;

17–33% – достаточно однородная;

35–40% – недостаточно однородная;

40–60% – это говорит о большом разбросе совокупности.

Т.е. считаем выборку однородной при $V\% \leq 33\%$

Стандартная ошибка
средней.

$$m = \frac{s}{\sqrt{n}}$$

Нижний квартиль Q_{25} – это значение случайной величины, ниже которого находится 25% выборки. В ранжированном ряду нижний квартиль находится под номером, определяемым по формуле: (округлить до ближайшего)

$$No = \frac{n + 1}{4}$$

Нижний квартиль Q_{25} – это значение случайной величины, ниже которого находится 25% выборки. В ранжированном ряду нижний квартиль находится под номером, определяемым по формуле: (округлить до ближайшего)

$$№ = \frac{n+1}{4}$$

№	рост мальчиков
1	167
2	168
3	170
4	170
5	170
6	172
7	176
8	178
9	182
10	186
11	188

$$№ = \frac{n+1}{4} \quad № = \frac{11+1}{4} = 3$$

→ Q25

Q25=170

n=11 (объем выборки)

№ порядковый номер

Верхний квартиль Q_{75} – это значение случайной величины, выше которого находится 25% выборки. В ранжированном ряду верхний квартиль находится под номером, определяемым по формуле:
$$№ = \frac{3(n+1)}{4}$$

№	рост мальчиков
1	167
2	168
3	170
4	170
5	170
6	172
7	176
8	178
9	182
10	186
11	188

$$№ = \frac{3(n+1)}{4} \quad № = \frac{3(11+1)}{4} = 9$$

Q75=182



n=11 (объем выборки)
№ порядковый номер

Пример

Рост мальчиков	186	178	167	170	168	172	182	176	170	188	170
Рост девочек	161	168	164	163	165	160	165	165	169	170	168

Educationproject - Microsoft Excel

Файл Главная Вставка Разметка стра Формулы Данные Рецензировани Вид Надстройки

Получение внешних данных Обновить все Подключения

Сортировка Фильтр Работа с данными Структура Анализ

Анализ данных

E5

	A	B
1	Рост мальчиков	Рост девочек
2	186	161
3	178	168
4	167	164
5	170	163
6	168	165
7	172	160
8	182	165
9	176	165
10	170	169
11	188	170
12	170	168
13		

Анализ данных

Инструменты анализа

- Однофакторный дисперсионный анализ
- Двухфакторный дисперсионный анализ с повторениями
- Двухфакторный дисперсионный анализ без повторений
- Корреляция
- Ковариация
- Описательная статистика**
- Экспоненциальное сглаживание
- Двухвыборочный F-тест для дисперсии
- Анализ Фурье
- Гистограмма

OK Отмена Справка

Нажать кнопку OK

Educationproject - Microsoft Excel

Файл Главная Вставка Разметка строки Формулы Данные Рецензирование Вид Настройки

Получение внешних данных Обновить все Подключения

Сортировка Фильтр Работа с данными Структура Анализ

	А	В
1	Рост мальчиков	Рост девочек
2	186	161
3	178	168
4	167	164
5	170	163
6	168	165
7	172	160
8	182	165
9	176	165
10	170	169
11	188	170
12	170	168
13		
14		
15		
16		
17		

Описательная статистика

Входные данные
Входной интервал:

Группирование:
 по столбцам
 по строкам

Метки в первой строке

Параметры вывода
 Выходной интервал:
 Новый рабочий лист:
 Новая рабочая книга

Итоговая статистика
 Уровень надежности: %
 К-ый наименьший:
 К-ый наибольший:

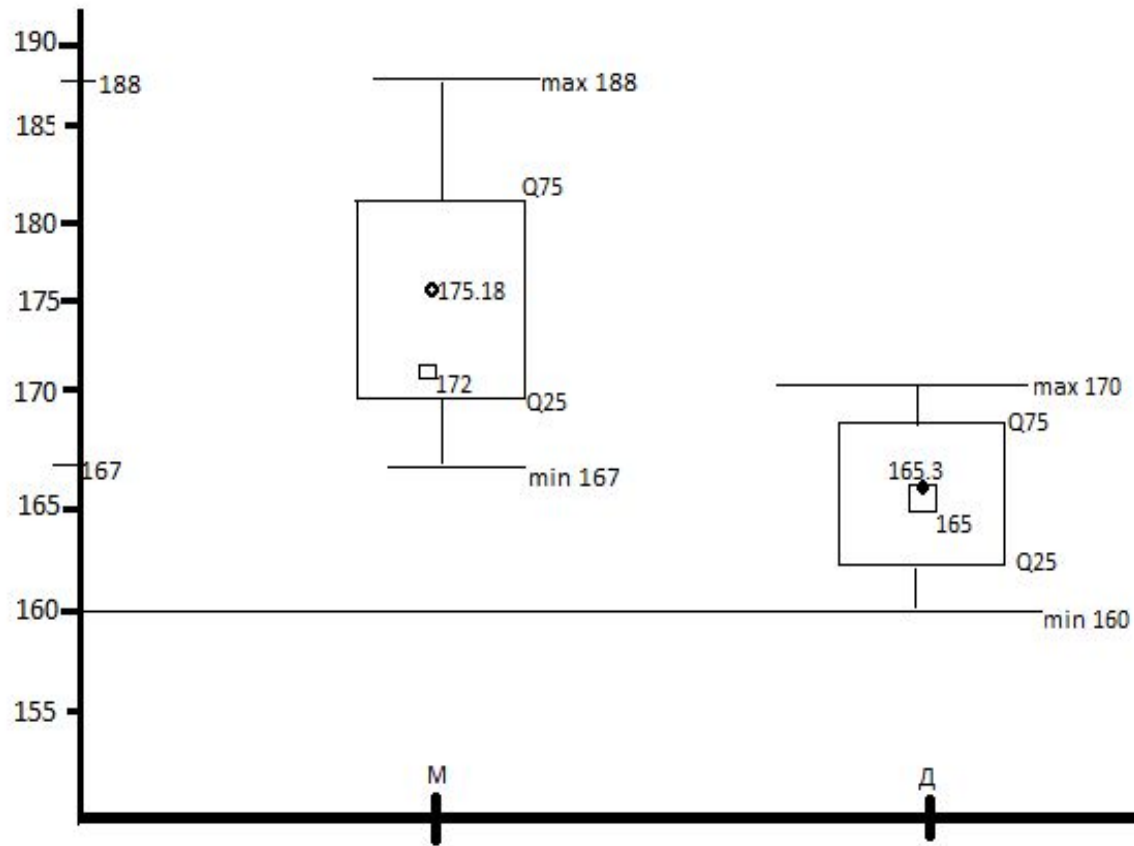
OK Отмена Справка

Нажать кнопку **OK**

Результат:

<i>рост мальчиков</i>		<i>рост девочек</i>	
Среднее	175,1818	Среднее	165,2727
Стандартная ошибка	2,22718	Стандартная ошибка	0,973194
Медиана	172	Медиана	165
Мода	170	Мода	165
Стандартное отклонение	7,38672	Стандартное отклонение	3,227721
Дисперсия выборки	54,56364	Дисперсия выборки	10,41818
Q25	170	Q25	163
Q75	182	Q75	168
Минимум	167	Минимум	160
Максимум	188	Максимум	170
Сумма	1927	Сумма	1818
Счет	11	Счет	11

График «ящик с усами»



Анализ результатов:

Мальчики более рослые, чем девочки – об этом свидетельствуют средние значения и положение выборок на числовой оси – выборка девочек расположена левее, т.е. в области меньших значений роста.

Дисперсия роста мальчиков, больше дисперсии роста девочек. Т.е. разброс значений роста у мальчиков больше, т.е. среди них встречаются низкорослые, среднего роста и высокие. На это указывают также коэффициент вариации и межквартильный размах.

У девочек группа более однородная, т.е. они не сильно различаются по росту.

У мальчиков наиболее часто встречается рост 170 см, об этом свидетельствует мода Mo . Половина мальчиков имеет рост меньше 172 см, на это указывает Me . Т.е. в целом мальчики невысокие.

У девочек наиболее часто встречается рост 165 см, об этом свидетельствует мода Mo . Половина девочек имеет рост меньше 165 см, на это указывает Me . Т.е. в целом девочки также