Общие сведения о надстройке «Пакет анализа» и статистических функциях MS Excel

Лекция 1

1	**Раздел 1. Методы описательной статистики**	**10**	**2**		**2**	**2**	**6**	
2	1. Общие сведения о надстройке «Пакет анализа» и статистических функциях MS Excel.	6	2				4	
3	1.2 Статистические функции, связанные с режимами «Гистограмма», «Ранг», «Описательная статистика»	4			2	2	2	
4	**Раздел 2. Методы дисперсионного анализа**	**12**	**2**		**4**	**4**	**6**	
5	Тема 2.1 Однофакторный дисперсионный анализ	6	2		2	2	2	
6	Тема 2.2 Двухфакторный дисперсионный анализ	6			2	2	4	
7	**Раздел 3. Статистические методы изучения взаимосвязей явлений и процессов**	**22**	**4**		**10**	**4**	**8**	
8	Тема 3.1 Корреляционный анализ	8	2		2		4	
9	Тема 3.2 Регрессионные модели	14	2		8	4	4	
10	**Раздел 4. Статистические методы изучения динамики явлений и процессов**	**28**	**6**		**12**	**4**	**10**	
11	Тема 4.1 Предварительный анализ временных рядов	6	2		2		2	
12	Тема 4.2. Модели кривых роста и тренд-сезонные модели	16	2		4	2	4	
13	Тема 4.3 Адаптивные модели прогнозирования	12	2		6	2	4	
14	Форма аттестационного испытания промежуточной аттестации	Зачет						

- **Тема 1.1. Общие сведения о надстройке «Пакет анализа» и статистических функциях MS Excel**

Обзор существующих программных средств обработки информации. Сводка и группировка данных с использованием электронных таблиц MS Excel. Работа с мастером функций. Методы компьютерной обработки статистических данных. Технология работы в режиме «Анализ данных». Применение пакета MS Excel при статистической обработке данных. Диалоговое окно режима «Описательная статистика».

- **Тема 1.2. Статистические функции, связанные с режимами «Гистограмма», «Ранг», «Описательная статистика»**

Функции описательной статистики: СРЗНАЧ, СРГАРМ, СРГЕОМ, МЕДИАНА, МОДА, КВАРТИЛЬ, ПЕРСЕНТИЛЬ, СТАНДОТКЛОН, ДИСП, КВАДРОТКЛ, СРОТКЛ, СТАНДОТКЛОНА, СТАНДОТКЛОНП, ЭКСЦЕСС, СКОС, МИН, МЕДИАНА, МАКС, МАКСА, НАИБОЛЬШИЙ, НАИМЕНЬШИЙ.

Гистограмма, алгоритм ее построения. Выборка. Технология формирования выборки из генеральной совокупности. Определение ранга числа в списке чисел с помощью функции ”РАНГ”.

Использование встроенных статистических функций, связанных с режимом с режимом «Описательная статистика», при анализе данных таможенной статистики.

Каждые два года объем данных, которыми обладает человечество, увеличивается в десять раз!

ЧТО НУЖНО ДЕЛАТЬ, ЧТОБЫ ДАННЫЕ ПРИНОСИЛИ ПОЛЬЗУ?

Анализ данных – процесс обнаружения в имеющихся данных ранее неизвестных, нетривиальных, практически полезных, доступных интерпретации данных, необходимых для принятия решений в различных сферах деятельности.

Важное замечание:

Анализ данных концентрируется на практическом применении статистических методов не только для того, чтобы делать выводы описательного характера (описательная и дескриптивная статистика), но и для того, чтобы предсказывать будущее изучаемых объектов (предиктивная аналитика) и давать рекомендации по принятию решений (предписывающая или прескриптивная аналитика).

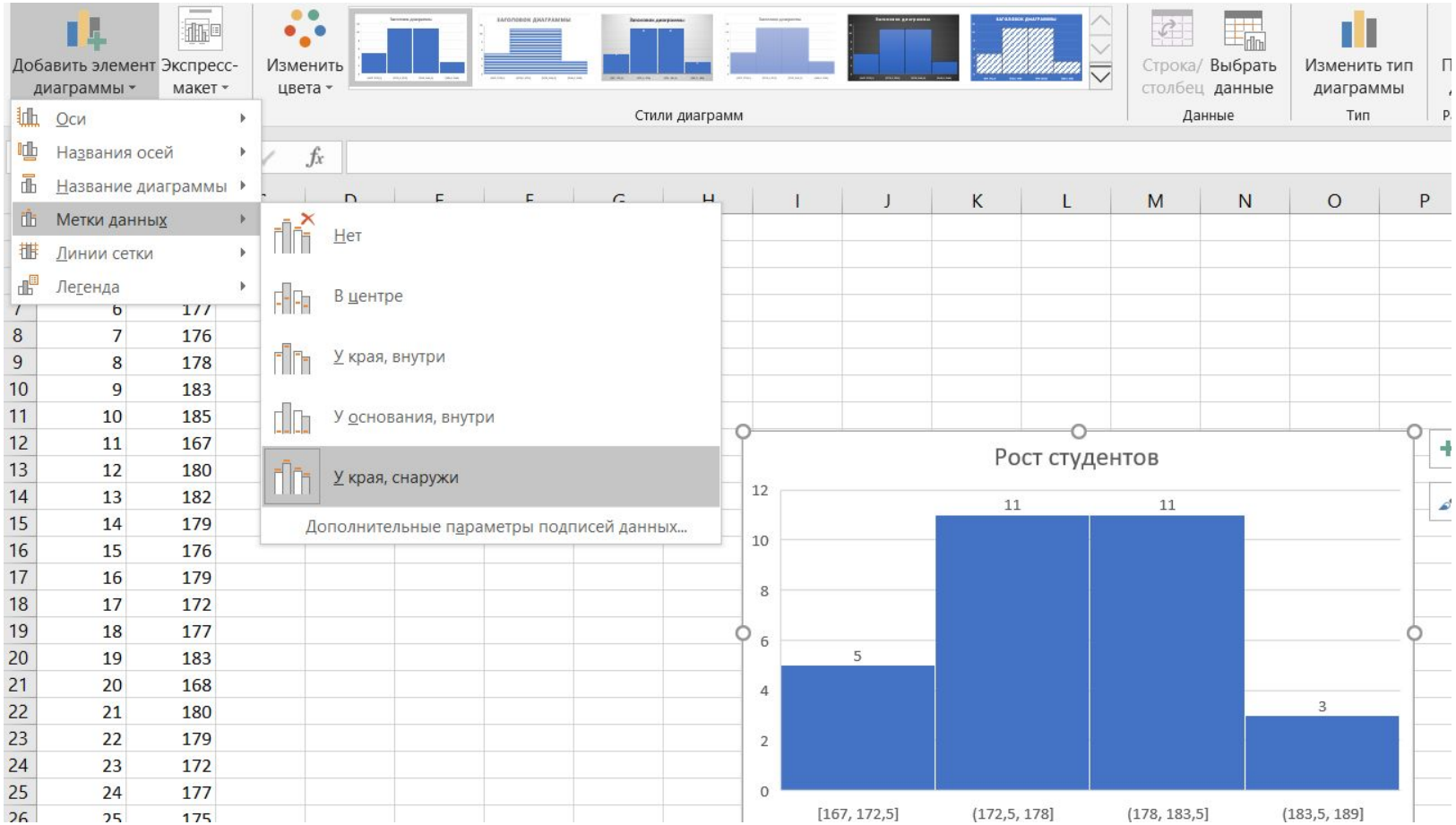
Описательная (дескриптивная) статистика

это раздел статистической науки, в рамках которого изучаются наиболее распространенные методы обработки статистических данных, включающие в себя их группировку, табличное и графическое представление, количественное описание с помощью основных статистических показателей (средние величины, меры рассеяния, характеристики формы распределения и другие).

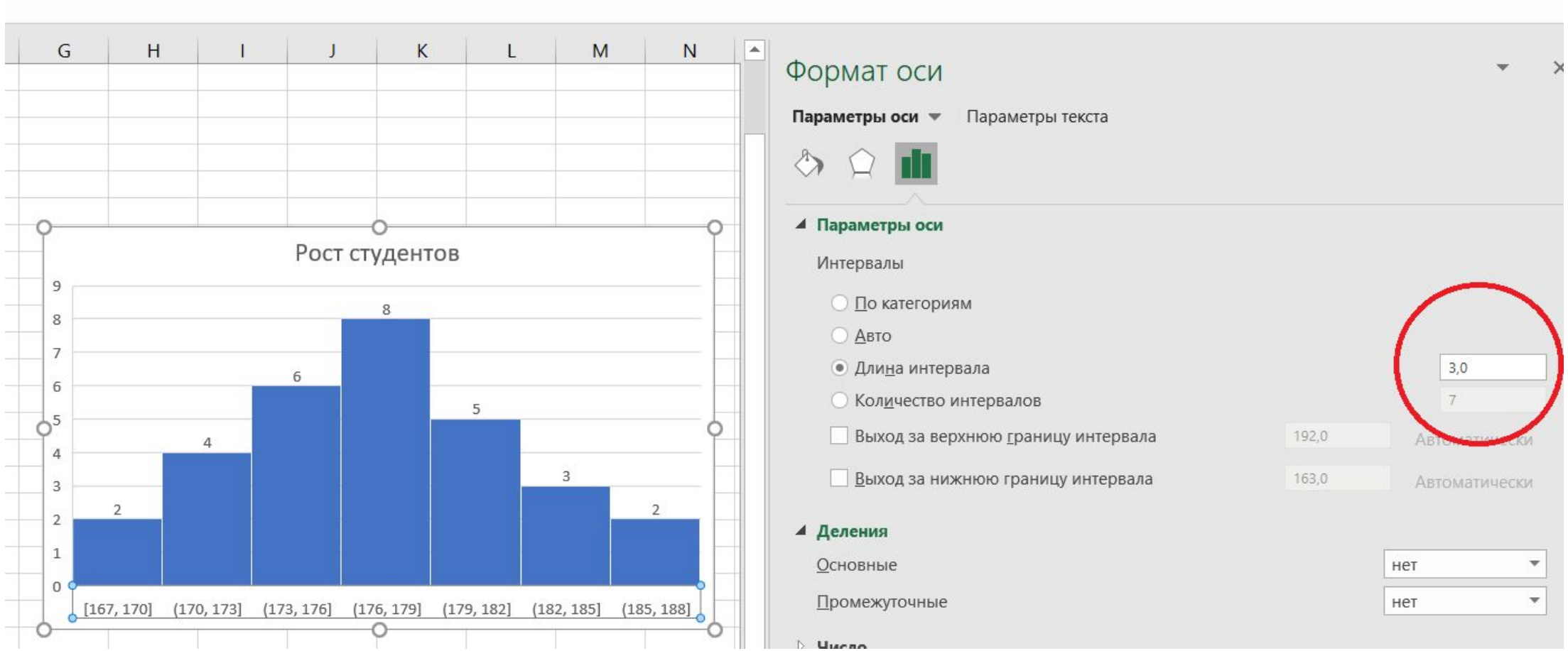
Данные о росте студентов учебной группы (см)

№ п/п	Рост	№ п/п	Рост	№ п/п	Рост
1	182	11	167	21	180
2	175	12	180	22	179
3	186	13	182	23	172
4	175	14	179	24	177
5	188	15	176	25	175
6	177	16	179	26	173
7	176	17	172	27	179
8	178	18	177	28	176
9	183	19	183	29	179,5
10	185	20	168	30	172

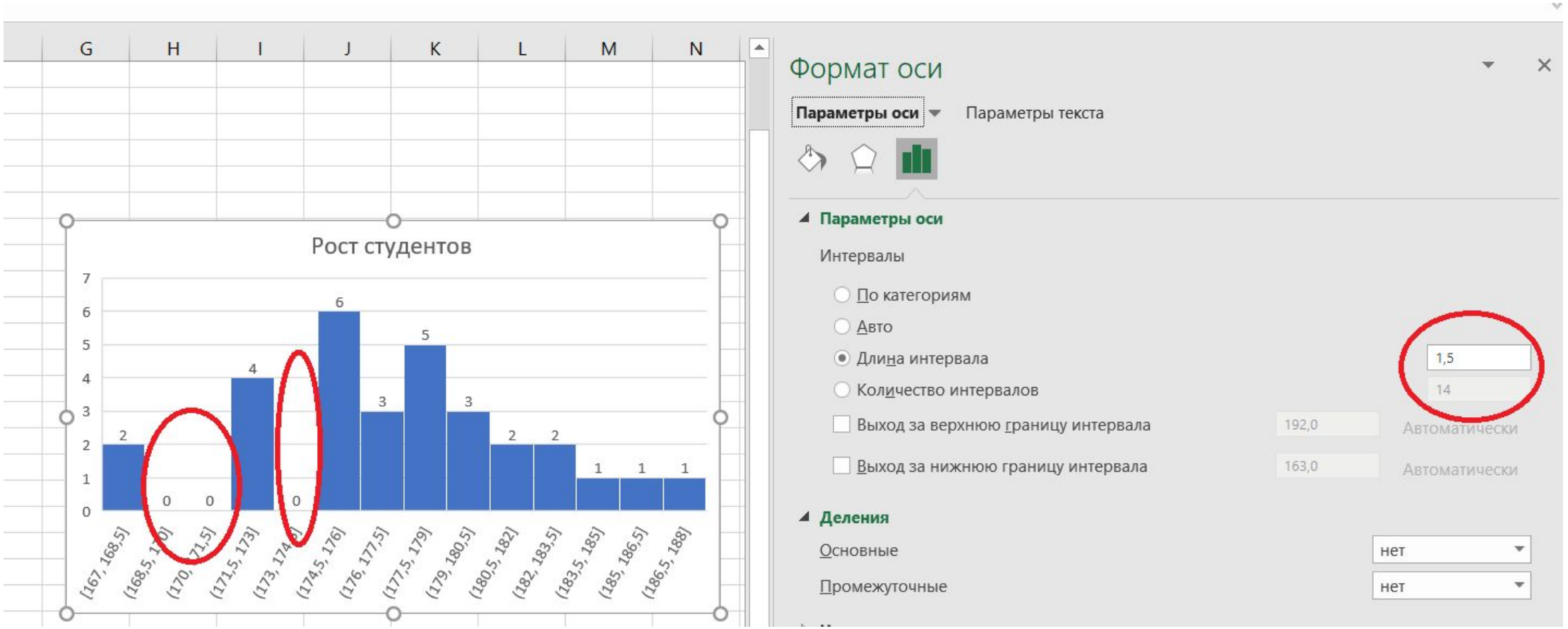
Вставка – статистическая гистограмма



Правая кнопка мыши на оси абсцисс – формат оси – длина интервала 3 см (справа).



Уменьшение длины интервала до 1,5 см



Вопрос:

Какие знаете правила, используемые для определения длины интервала группирования?

Формула Стерджеса

Правило Скотта

Правило Фридмана-Диаконисса

Что же делать?

На практике лучше всего попробовать несколько вариантов и выбрать из них тот, при котором гистограмма выглядит наиболее «гладкой»

Альтернативный вариант:

Данные-анализ данных-гистограмма

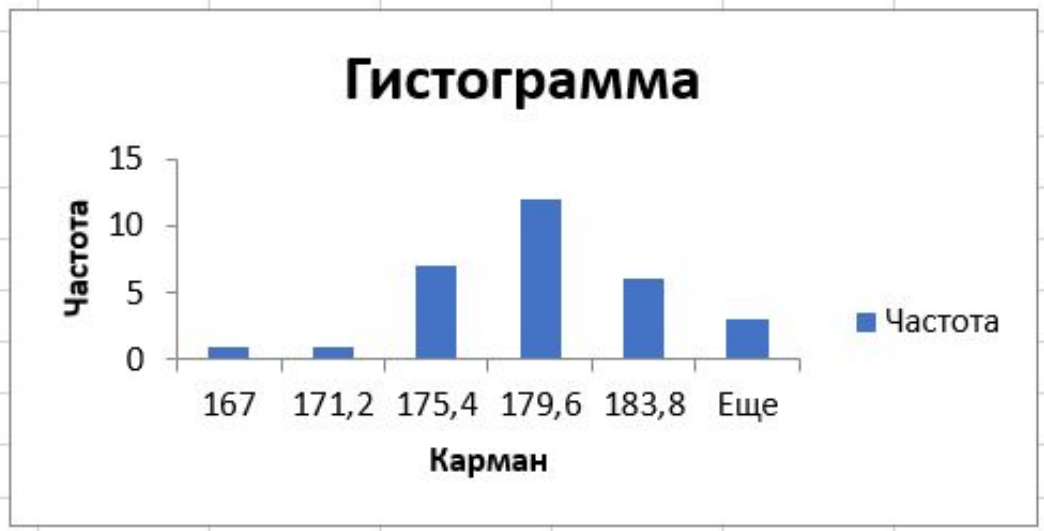
Гистограмма. Используется для вычисления выборочных и интегральных частот попадания данных в указанные интервалы значений. При этом рассчитываются числа попаданий для заданного диапазона ячеек.

Элементы диалогового окна «Гистограмма» включают в себя:

- Входной диапазон: ссылка на диапазон, содержащий анализируемые данные.
- Интервал карманов (необязательный): диапазон ячеек и необязательный набор граничных значений, определяющих отрезки (карманы). Эти значения должны быть введены в возрастающем порядке. В Microsoft Excel вычисляется число попаданий данных между текущим началом отрезка и соседним большим по порядку, если такой есть. Включаются значения на нижней границе отрезка и не включаются значения на верхней границе.
- Парето (отсортированная диаграмма): если установлен флажок, то данные будут представлены в порядке убывания частоты. Если флажок снят, то данные в выходном диапазоне выводятся в порядке возрастания отрезков.
- Интегральный процент: если установлен флажок, то будут вычислены и графически отмечены накопленные частости.
- Вывод графика: если установлен флажок, то автоматически создается встроенная диаграмма - гистограмма.

176
79,5
172

Карман	*Частота*
167	1
171,2	1
175,4	7
179,6	12
183,8	6
Еще	3



Описательная статистика

дает ответы на вопросы о том, что общего есть в анализируемых данных и какие в них есть различия.

другими словами, это расчет определенных статистических показателей для:

-измерения центра распределения

-измерения разброса данных

Описательная статистика в MS EXCEL

- Следует отметить, что в MS Excel реализованы статистические функции, вычисляемые только ***по простым формулам***.
- В частности, функции СРЗНАЧ(), СРГАРМ(), СРГЕОМ() возвращают среднее значение (среднее арифметическое, гармоническое, геометрическое, соответственно) аргументов. Например, если диапазон C5:C10 содержит числа, формула **=СРЗНАЧ(C5:C10)** возвращает **среднюю арифметическую простую** этих чисел.
- Структурные средние медиану и моду вычисляют с помощью функций МЕДИАНА(), МОДА(), аргументами которых должны являться числовые массивы. Функции КВАРТИЛЬ(), ПРОЦЕНТИЛЬ () возвращает k-ю квартиль и процентиль для значений диапазона, соответственно.

Описательная статистика в MS EXCEL

- Функции ДИСП.Г(), ДИСП.В(), СРОТКЛ() возвращают генеральную дисперсию, исправленную дисперсию и среднее абсолютное линейное отклонение значений аргументов, соответственно. Логические значения и текст игнорируются. Функция КВАДРОТКЛ() возвращает сумму квадратов отклонений точек данных от их среднего. Для вычисления выборочного и исправленного среднеквадратического отклонения применяются функции СТАНДОТКЛОН.Г() и СТАНДОТКЛОН.В(), соответственно. Аргументами функций должны являться числовые массивы, логические значения и текст игнорируются.
- Несмещенные статистические оценки коэффициентов асимметрии и эксцесса возвращают функции ЭКСЦЕСС() и СКОС(), соответственно.
- Минимальное и максимальное значения из числового массива позволяют выделить функции МИН() и МАКС(), соответственно. Наряду с указанными функциями имеется возможность определить k-ое по величине значение из множества данных с помощью функции НАИБОЛЬШИЙ(), аргументами которой являются массив данных и значение k. Аналогично определяется функция НАИМЕНЬШИЙ().

ОПИСАТЕЛЬНАЯ СТАТИСТИКА В НАДСТРОЙКЕ «АНАЛИЗ ДАННЫХ»

Описательная статистика.

Это средство анализа служит для создания одномерного статистического отчета, содержащего информацию о центральной тенденции и изменчивости входных данных

Элементы диалогового окна «Описательная статистика»:

- Входной диапазон: ссылка на диапазон, содержащий анализируемые данные, состоящие не менее чем из двух смежных диапазонов данных, данные в которых расположены по строкам или столбцам.
- Группирование: в зависимости от расположения данных во входном диапазоне, следует установить переключатель в положение «*По столбцам*» или «*По строкам*».

Элементы диалогового окна «Описательная статистика» (продолжение):

- Уровень надежности: для того чтобы в выходную таблицу включить строку для уровня надежности, необходимо установить соответствующий флажок и ввести в поле требуемое значение надежности интервальных оценок.
- Выходной диапазон: необходимо ввести ссылку на левую верхнюю ячейку выходного диапазона. Этот инструмент анализа выводит два столбца сведений для каждого набора данных, в которых левый столбец содержит названия описательных статистик; а правый – их значения.

Элементы диалогового окна «Описательная статистика» (окончание):

Итоговая статистика: необходимо установить флажок для вывода значений следующих описательных статистик:

- Среднее, Стандартная ошибка (среднего),
- Медиана, Мода,
- Стандартное отклонение, Дисперсия выборки,
- Эксцесс, Асимметричность,
- Интервал,
- Минимум, Максимум,
- Сумма, Счет,
- Уровень надежности.

Результаты расчета описательной статистики по данным о росте студентов

Столбец1	
Среднее	177,6833
Стандартная ошибка	0,898429
Медиана	177,5
Мода	179
Стандартное отклонение	4,920897
Дисперсия выборки	24,21523
Эксцесс	0,059917
Асимметричность	-0,08915
Интервал	21
Минимум	167
Максимум	188
Сумма	5330,5
Счет	30