

Цель и задачи дисциплины

Биометрия (от био... и...метрия), **раздел биологии**, содержанием которого являются

- планирование
- и обработка результатов количественных экспериментов и наблюдений

методами математической статистики.

Биологической статистикой или **биометрией** называется область научного знания, охватывающая

- классификацию,
 - систематизацию
 - обработку экспериментальных данных
- в биологии, медицине и сельском хозяйстве **методами математической статистики.**

С формальной точки зрения
биометрия представляет
«**совокупность математических**
методов, применяемых в биологии».

Теория вероятностей

Изучает законы поведения случайных событий и случайных переменных величин

Разделы математики

Математическая статистика

Занимается разработкой теории выборочного метода

Биометрия

Анализирует конкретные факты с помощью **методов**

□ **математической статистики**

□ **теории вероятностей,**

составляющих в совокупности то, что называют

статистическим анализом.

наука прикладная

Имеет дело с конкретными фактами

- **Математическая биология** подходит к решению биологических проблем **дедуктивно**, выдвигая на первый план математические модели с последующей проверкой их опытом.
- **Биометрия** опирается на индуктивный метод, отправляясь от конкретных фактов, которые она анализирует методами математической статистики и теории вероятностей.

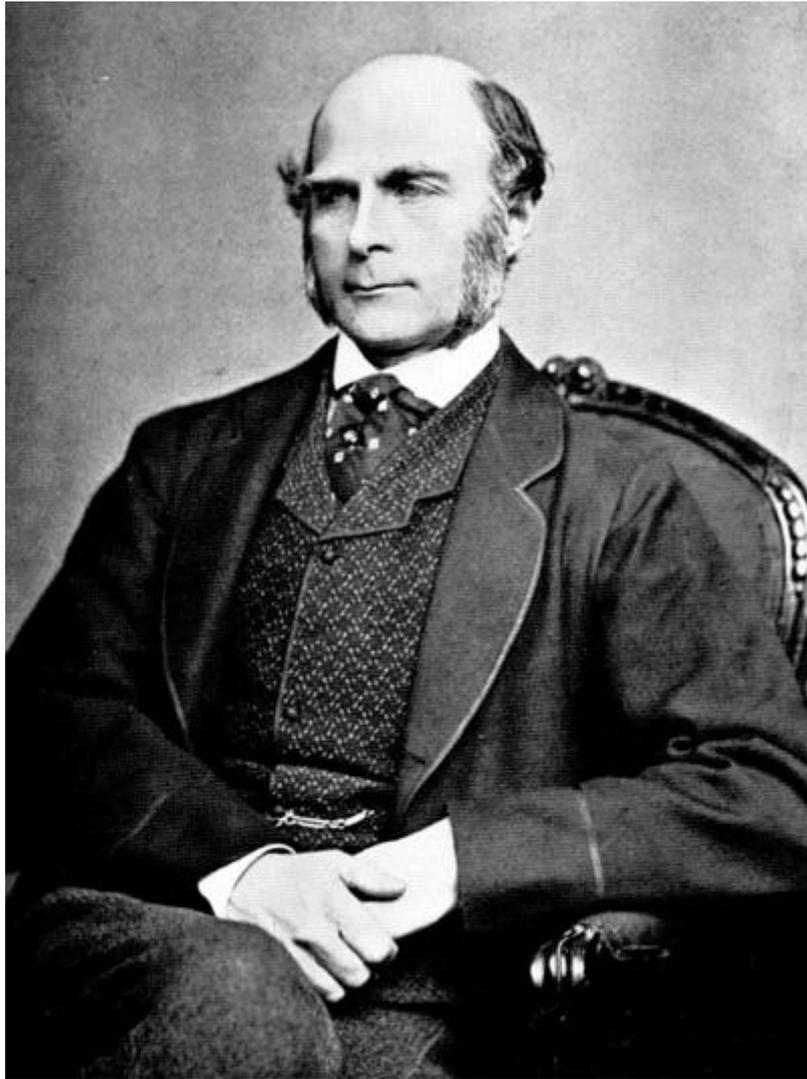
- ✓ Биометрия как самостоятельная наука возникла в XIX веке.
- ✓ Быстрое развитие теоретической и прикладной математики в XVII веке привело к важным результатам.
- ✓ В середине XVII века, независимо одна от другой, зародились еще две ветви точных наук:
 - теория вероятностей
 - математическая статистика.

Дальнейшее развитие биометрия получила главным образом в трудах английских ученых

- Ф. Гальтона (1822-1911)
- К. Пирсона (1857-1936),

основавших известную школу биометриков и создавших математический аппарат этой науки.

Френсис Гальтон



1822-1911

- Заложил основы новой науки и дал ей имя.
- Впервые применил **статистический метод А. Кетле** к решению кардинальной проблемы **наследственности и изменчивости организмов.**
- Разрабатывает методику регрессного и корреляционного анализа, вошедшую в золотой фонд биометрии.

НАСЛѢДСТВЕННОСТЬ ТАЛАНТА

ЕЯ ЗАКОНЫ И ПОСЛѢДСТВІЯ.

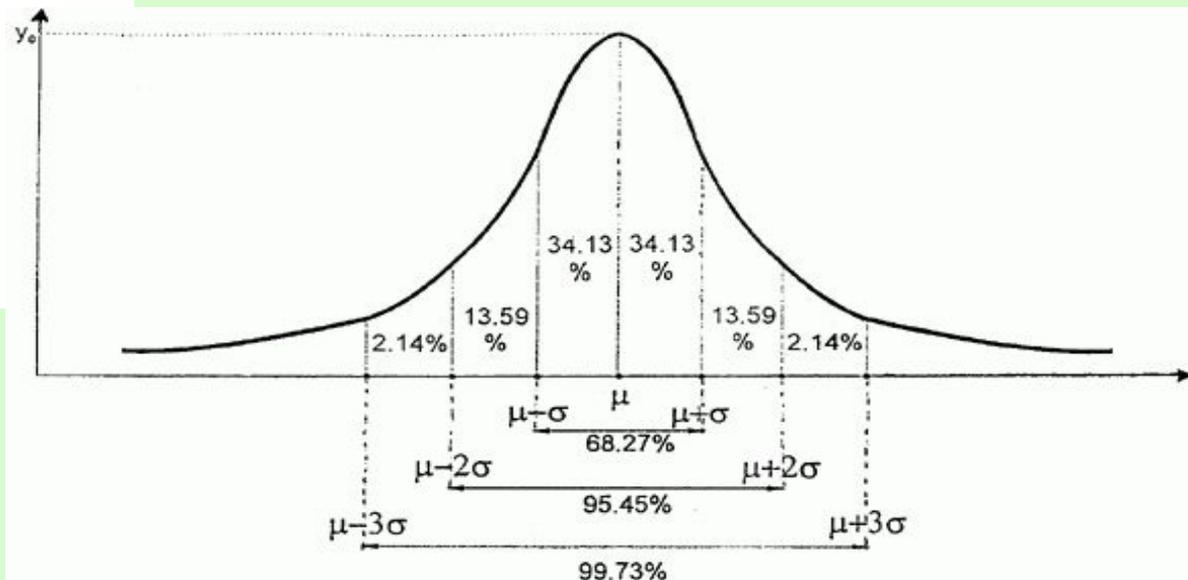
Фрэнсиоа Гальтона.

ПЕРЕВОДЪ СЪ АНГЛІЙСКАГО.

С.-ПЕТЕРБУРГЪ
Изданіе редакціи журнала «Знаніе»
1875

Подтвердил вывод А. Кетле

- ❖ не только физические признаки, но и умственные способности человека распределяются по нормальному закону, описываемому формулой Гаусса-Лапласа.



Карл Пирсон



1857 – 1936

1893 г. - понятие среднего квадратического отклонения
- коэффициента вариации

1898 г. – разрабатывает основы множественной регрессии

1903 г. - разработал основы теории сопряженности признаков

1905 г. - опубликовал основы нелинейной корреляции и регрессии

- обоснование метода хи-квадрат

- развил учение о типах кривых распределения

- разработал методику построения множественной регрессии

С именем **К. Пирсона (1857 – 1936)** связано:

- обоснование метода хи-квадрат, получившего широкое распространение в биометрии;
- разработал учение о типах кривых распределения;
- разработал теорию линейной и нелинейной корреляции;
- разработал методiku построения множественной регрессии.

Рональд Фишер



1890-1962

- ❑ Создана методология современной биометрии
- ❑ Впервые показал, что планирование экспериментов и наблюдений и обработка их результатов — две неразрывно связанные задачи статистического анализа.

Р. Фишером были разработаны:

- теория выборочных распределений;
- методы **дисперсионного** и дискриминантного анализа;
- теории планирования экспериментов;
- метод максимального правдоподобия,

и многое другое, что составляет основу современной прикладной статистики и математической генетики.

Такое сочетание статистики и генетики не является чем-то редким.

- **Распространение биометрических идей и методов русскими учёными**

Четвериков Сергей Сергеевич

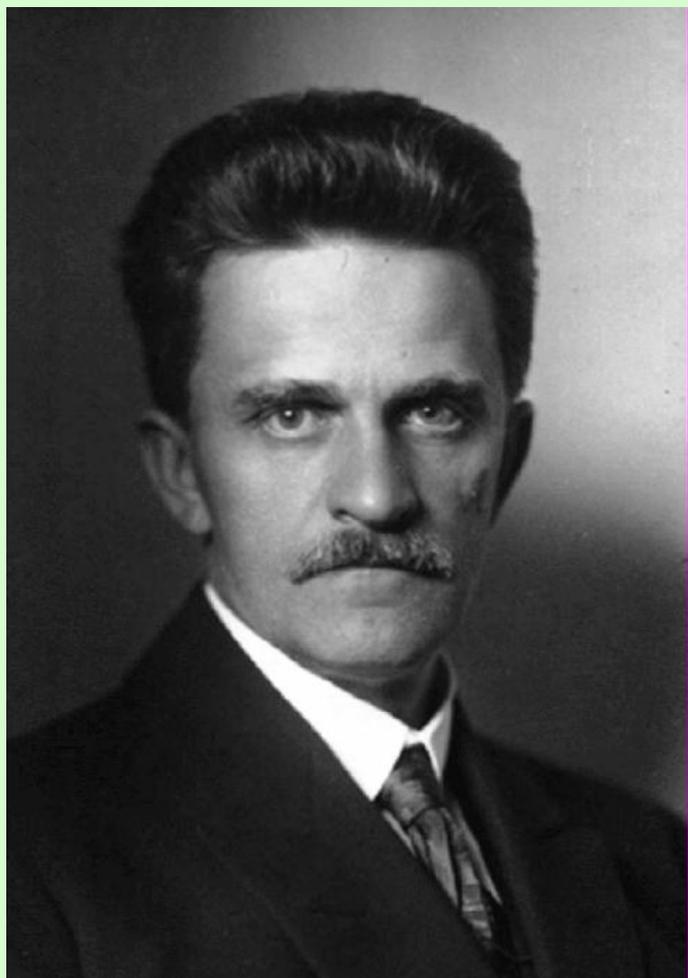


Первым, кто еще в 1919 г. начал читать студентам МГУ курс биометрии с основами генетики, был С.С. Четвериков.

В 1924 г. в МГУ уже самостоятельный курс биометрии.

Филипченко Юрий Александрович -

советский биолог и генетик, известный своей педагогической и научно-организаторской деятельностью



1882-1930

Его научные интересы охватывали: генетику качественных и количественных признаков, включая наследование таланта у человека, евгенику, генетические основы эволюции. Он предложил понятия «микроэволюция» и «макроэволюция».

Организовал при ЛГУ первую в России и вторую в мире кафедру генетики.

Основатель ленинградской школы биометриков.

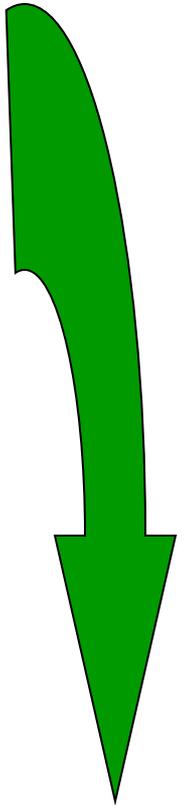
Статистические методы обработки экспериментальных данных используют для:

- характеристики биологических явлений
-
- технологических процессов производства
- переработки сельскохозяйственной продукции.

Необходимость статистической обработки данных

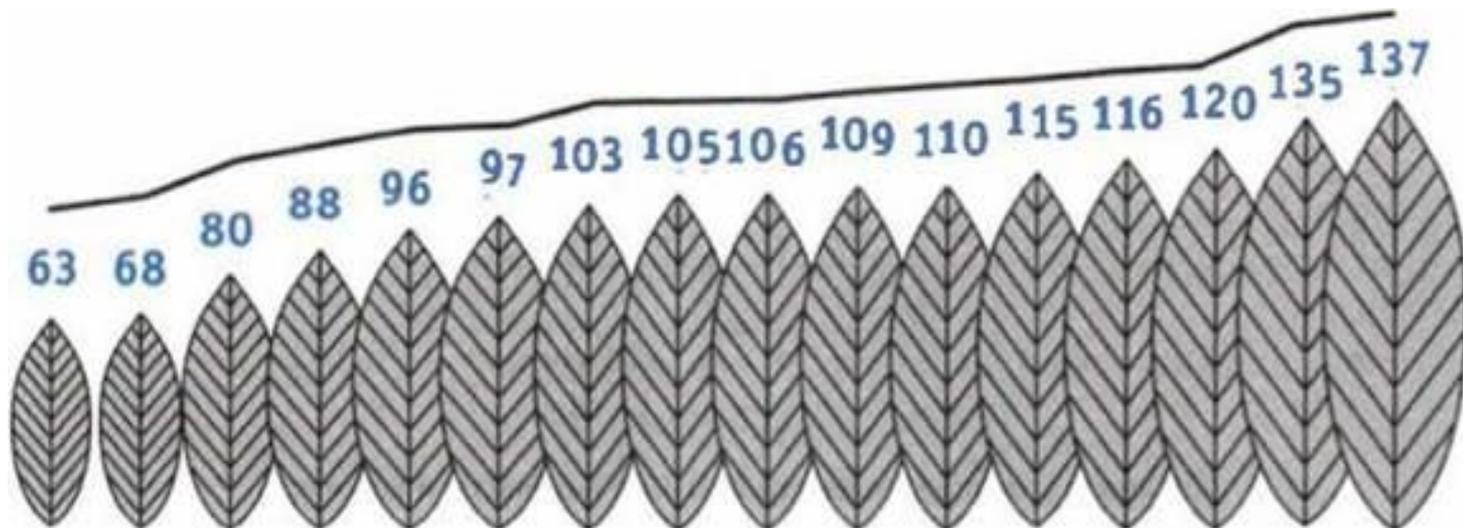
- При переходе от описательного метода анализа биологических явлений к экспериментальному.
- Признаки и свойства характеризуются количественными вариациями частоты встречаемости или степени проявления.

Каковы возможные пределы случайных колебаний изучаемой величины и являются ли наблюдаемые разницы между вариантами опыта случайными или достоверными.



Необходимость использования статистических расчетов экспериментальных данных

1. Все биологические объекты обладают двумя противоположными свойствами: наследственностью и изменчивостью. Широкая амплитуда изменчивости признаков у различных объектов требует усреднения данных, оценить границы изменчивости и силу связи между признаками.



- 2. Все биологические явления и свойства подчиняются статистическим закономерностям, характерным не отдельным объектам, а целым совокупностям объектов.

Статистические методы анализа биологических объектов позволяют

- определить средние величины изучаемого признака (среднее арифметическое – \bar{x} ; мода – M_o ; медиана – M_e);
- установить характер и тип распределения объектов с разными параметрами признака (нормальное, биномиальное, Пуассона и др.)
- выявить изменчивость признака с помощью среднего квадратического (или стандартного) отклонения (σ), дисперсии (σ^2), дисперсии (S), коэффициента изменчивости (C_v);

Статистические методы анализа биологических объектов позволяют

- оценить значимость различия показателей в разных совокупностях (использования критерия достоверности Стьюдента, Фишера, метода χ^2 и других непараметрических методов);
- определить величину и направление связи между переменными величинами признаков объектов совокупности (коэффициенты корреляции (r) и регрессии (R); ранговый коэффициент Спирмена (r_s));
- изучить степень влияния того или иного фактора на изменчивость анализируемого признака (дисперсионный анализ) и прогнозировать показатели-отклики при заданных значениях воздействующих факторов.

Предметом биометрической генетики служит группа биологических объектов, которая составляет **совокупность**.

Всякое множество идентифицируемых объектов, отличающихся друг от друга незначительно по конкретному признаку, но сохраняющих сходство по некоторым существенным характеристикам, называется **совокупностью**.

Совокупностями могут являться сорта, виды растений, делянка или поле растений.

Совокупность состоит из **единиц** или **членов**. Число единиц, входящих в совокупность, называется **объектом совокупности** и обозначается буквой **n**. Единица совокупности характеризуется определенными признаками.

Величину изучаемого признака для отдельной единицы совокупности называют **вариантой** и обозначают **x1, x2, x3, ..., xi**, а в общем виде **xi**, где индекс **i** означает порядковый номер варианты.

Например, при изучении урожая пшеницы определенного сорта с 1 га получены следующие данные 25, 27, 28,5 ц. Эти величины и будут вариантами, т.е. $x_1=25$, $x_2=27$, $x_3=28,5$.

Различия между отдельными вариантами называются **изменчивостью** или **вариацией**.

Различают **количественную** и **качественную** изменчивость.

Количественная изменчивость бывает двух типов:

- прерывная (дискретная)
- непрерывная.

Если же различия определяются целыми числами, то это будет **прерывная (дискретная)** изменчивость. Так, число колосков выражается целым числом 11, 12, 13 и т.д.

При **непрерывной изменчивости** между вариантами нет четких границ и переходов, различия между вариантами определяются точностью измерений (длина колоса, вес зерна, колоса).

Качественные признаки – это признаки, которые описываются словами и отражают качественные характеристики объекта.

Например: окраска семян гороха (желтая и зеленая).

Большинство качественных признаков характеризуется альтернативной изменчивостью. Например: здоровые и больные растения.

2. Генеральная и выборочная совокупность

- **Генеральная совокупность** – все множество имеющихся объектов.
Под **генеральной совокупностью** понимается все множество возможных объектов, характеризующихся определенным признаком .
- **Выборка** (выборочная совокупность) – набор объектов, случайно отобранных из генеральной совокупности.
- **Репрезентативная выборка** (представительной).
Если в выборку входит до 30 членов, она называется малой ($n < 30$), а свыше 30 – большой ($n > 30$).
Из выборки можно выбрать еще меньшую выборку.
- Объем **генеральной совокупности** N и **объем выборки** n – число объектов в рассматриваемой совокупности.

Исходные данные

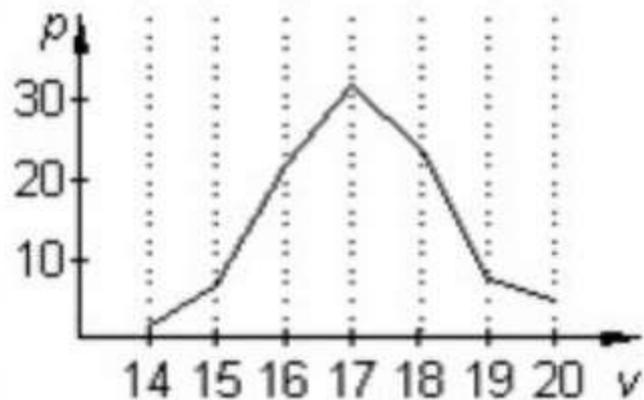
45	49	50	51	52	53	54	54	63	60
49	50	52	53	62	61	59	57	56	55
50	51	52	53	59	57	56	55	55	54
51	52	53	54	55	55	54	55	54	53
52	53	58	56	55	54	53	52	51	54
54	60	57	56	55	54	53	51	54	49
54	59	55	54	53	52	54	47	54	55
59	55	60	52	54	48	58	57	54	58
52	48	50	54	56	54	57	51	46	54
53	55	56	55	47	50	52	51	55	55

- Последовательность вариантов, записанных в порядке возрастания, называют **вариационным рядом**, а перечень вариантов и соответствующих им частот или относительных частот – **статистическим рядом**.

Вариационный ряд

- Например, если взять 100 колосьев пшеницы ($n = 100$), подсчитать число колосков в колосе (v) и число колосьев с данным количеством колосков, то вариационный ряд будет выглядеть следующим образом.

Вариант a (v)	14	15	16	17	18	19	20
Частота встречаемости (p)	2	7	22	32	24	8	5



Вариационная кривая

На основании вариационного ряда строится вариационная кривая - графическое отображение частоты встречаемости каждой варианты. Среднее значение признака встречается чаще, а вариации, значительно отличающиеся от него, - реже. Это называется «нормальным распределением». Кривая на графике бывает, как правило, симметричной.

Статистические показатели для характеристики совокупности

Иными словами:

Статистический показатель – это обобщающая характеристика какого – то свойства совокупности, группы.

Этим он отличается от индивидуальных значений, которые, называются признаками.

✓ **Например**, средняя продолжительность ожидаемой жизни родившегося поколения людей в стране – статистический показатель.

Продолжительность жизни конкретного человека – признак.

Основные параметры для характеристики совокупности

- Средние величины
- Показатели изменчивости признака

Статистические показатели для характеристики совокупности

□ Средние величины

- Среднее арифметическое (\bar{x}) показывает, какое значение признака наиболее характерно в целом для данной совокупности.
- Является обобщенной характеристикой совокупности.
- Она используется как для характеристики отдельных выборочных совокупностей (например, сорта) по какому-либо признаку, так и для сравнения их между собой.

3. Сущность средних величин

Средняя арифметическая – наиболее распространенный вид средних величин. Она используется тогда, когда известны отдельные значения осредняемого признака и общее число единиц совокупности

Средняя арифметическая простая имеет вид:

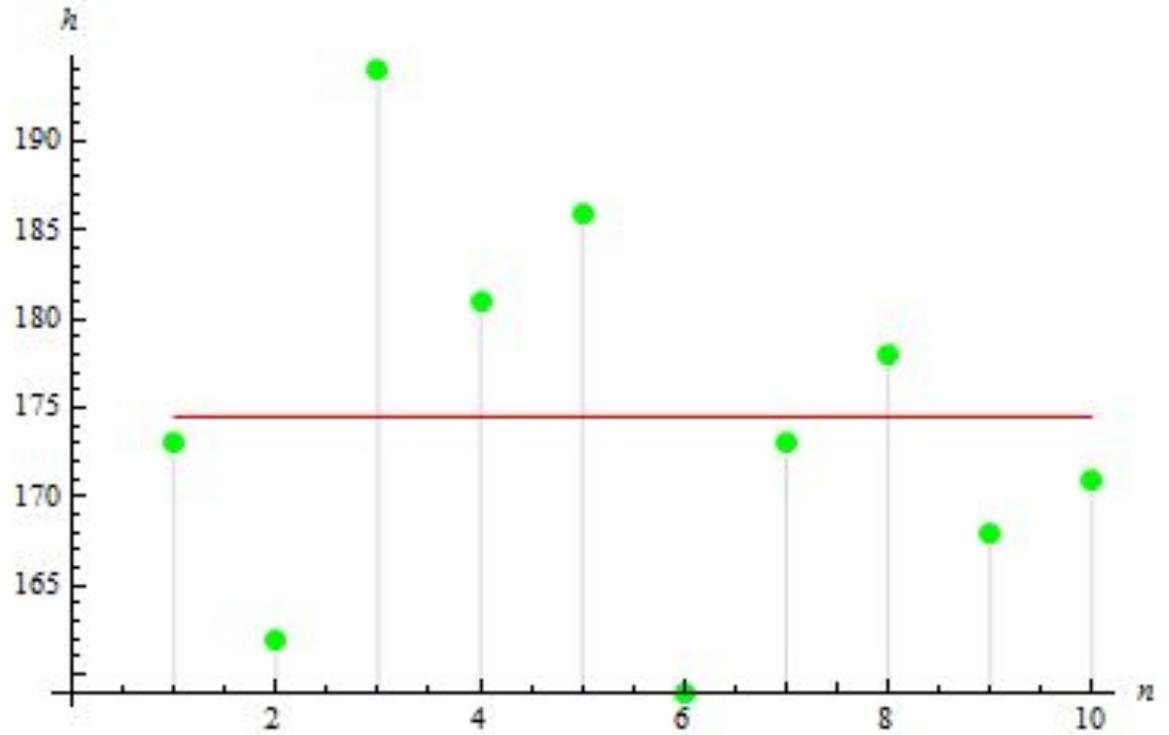
$$\bar{x} = \frac{\sum x}{n}$$

где $\sum x$ – сумма всех показателей, подлежащих
изучению

n – количество показателей, подлежащих
изучению.

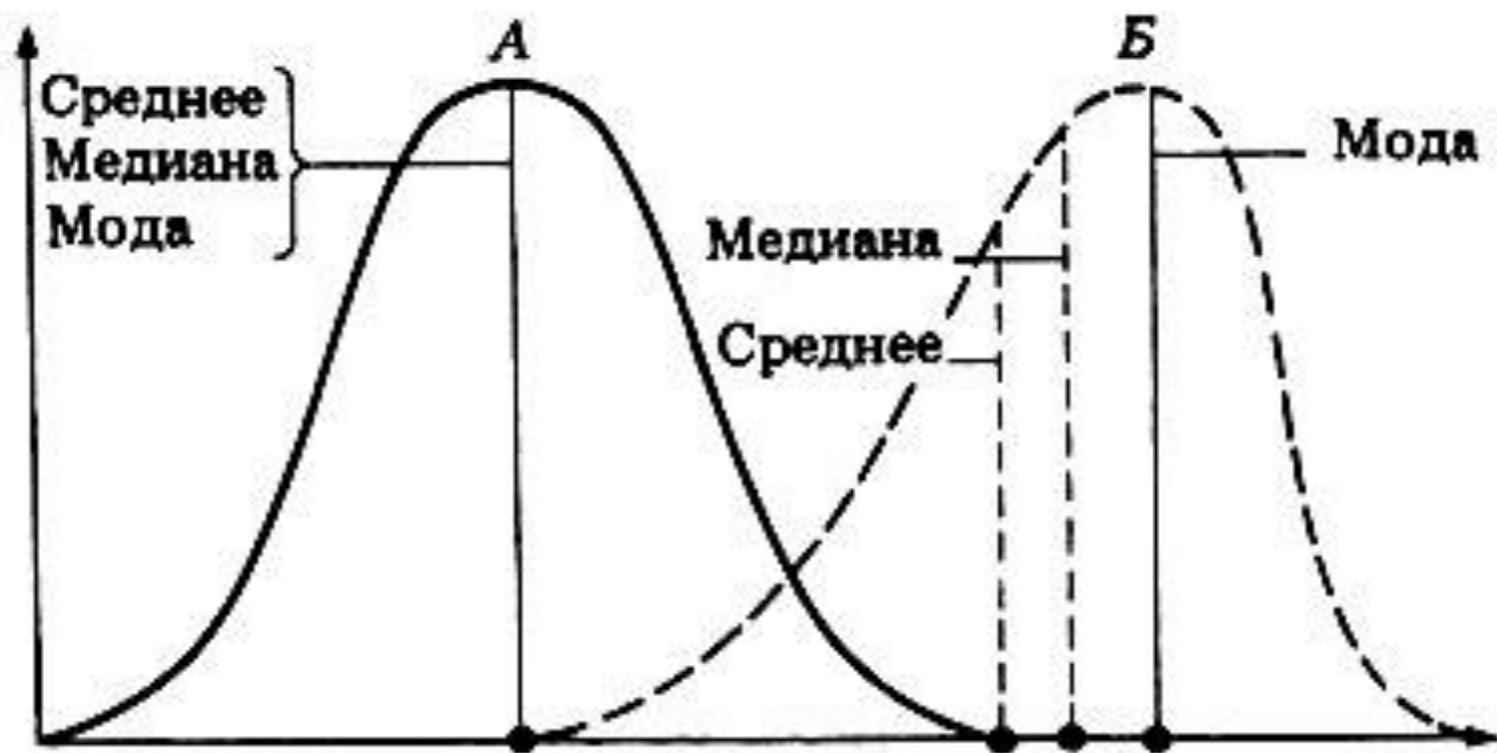
	1	2	3	4	5	6	7	8	9	10
Рост	173	162	194	181	186	159	173	178	168	171

174.5 cm.



- Мода (Mo) – наиболее часто встречающаяся варианта в совокупности.
- Медиана (Me) – варианта, расположенная в середине (центре) ряда и делящая его на две равные части. В ряду с чётным числом наблюдений находятся две варианты, тогда за Me принимается их полусумма.





Если выборочная совокупность имеет одну M_0 , то распределение объектов (наблюдений) в такой выборке будет называться унимодальным.

В случае наличия двух и более M_0 речь идет о би- или полимодальном распределении объектов (наблюдений) в совокупности. Полимодальный ряд распределения свидетельствует о неоднородности выборочной совокупности, т.е. наблюдается объединение качественно различных совокупностей.



Определение вариации

- Вариацией называется изменчивость значений признака у единиц статистической совокупности

Необходимость измерения вариации



- Средняя величина характеризует совокупность по изучаемому признаку, такой характеристики совокупности будет достаточно, если разброс индивидуальных значений невелик
- Когда ряд характеризуется значительным рассеиванием индивидуальных значений, то применение средней величины ограничено

Показатели изменчивости

Для характеристики разнообразия признаков в совокупности служат:

- ЛИМИТЫ
- варианса
- коэффициент вариации
- среднее квадратическое отклонение.

Абсолютные показатели

ЛИМИТЫ

Размах вариации

$$R = x_{\max} - x_{\min}$$

Недостаток РВ: он учитывает только крайние значения и не учитывает промежуточные значения

- **Лимиты (lim)** - это максимальное и минимальное значения признака в совокупности.
- Чем больше разность между максимальной (max) и минимальной (min) вариантой, тем выше изменчивость признака. Однако при одинаковых лимитах изменчивость в сравниваемых группах может различаться, так как лимиты не учитывают распределения отдельных вариантов в совокупности.

- **Дисперсия (S)** – сумма квадратов отклонений каждой варианты от среднего арифметического.

Дисперсия

Варианса

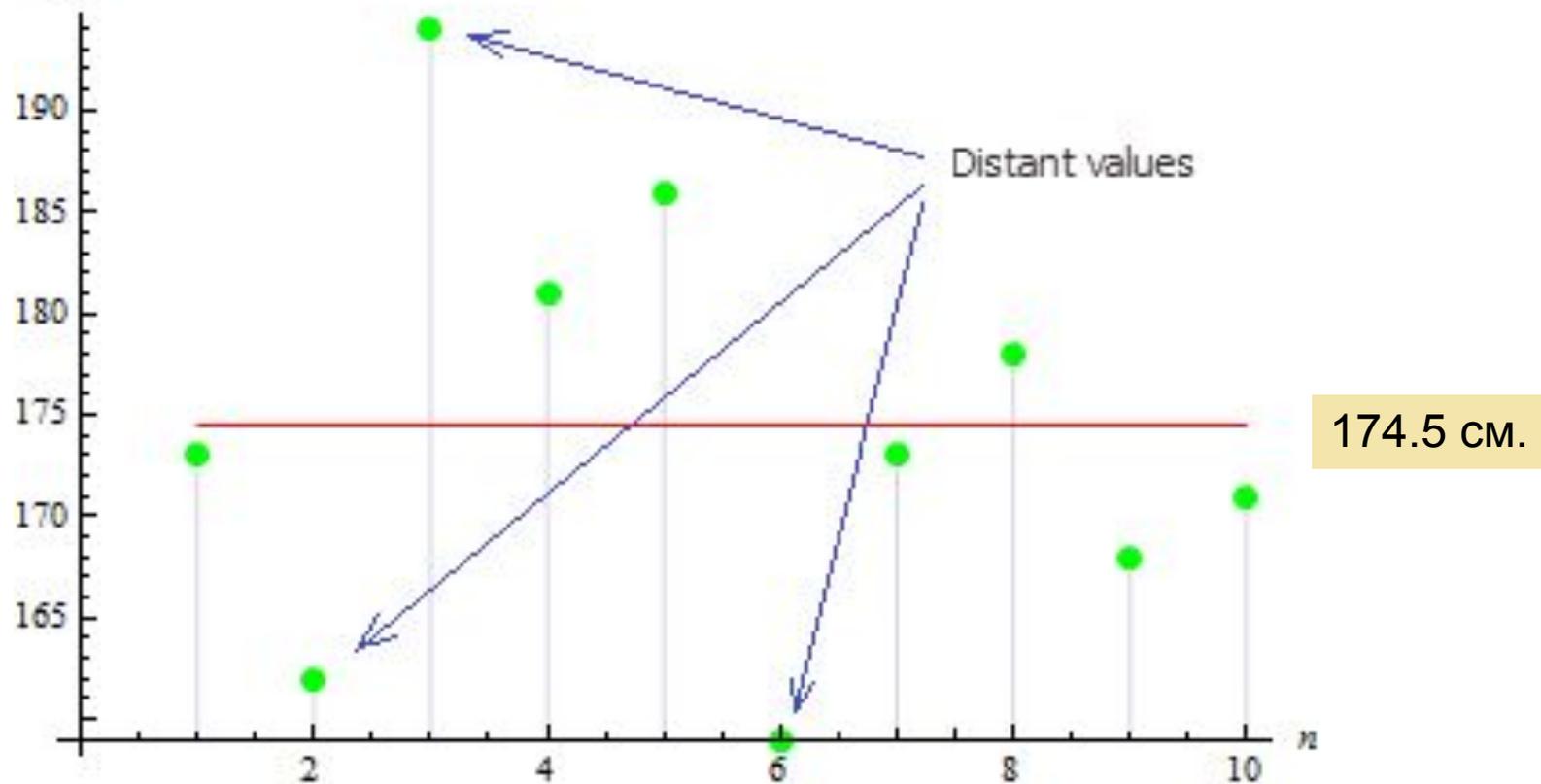
а) для несгруппированных данных

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$$

б) для сгруппированных данных

$$\sigma^2 = \frac{\sum (x - \bar{x})^2 \cdot f}{\sum f}$$

1	2	3	4	5	6	7	8	9	10
173	162	194	181	186	159	173	178	168	171



Значения далеко отстоят от мат. ожидания, что и приводит к большому значению дисперсии. 113.611.

Вариансой, или средним квадратом, называют сумму квадратов центральных отклонений, деленную на число степеней свободы.

- df - число степеней свободы, т. е. количество всех вариантов совокупности, уменьшенных на единицу ($df = n - 1$). Для выборки из 100 особей ($n = 100$) число степеней свободы равно 99 ($n - 1 = 100 - 1 = 99$).
- Варианса характеризует степень разнообразия величин, собранных в одну группу. Если выборка составлена из отдельных измерений признака, варианса характеризует разнообразие вариантов этой группы по данному признаку.
- Если группа составлена из средних величин для выборок, взятых из одной генеральной совокупности, то s^2 характеризует получившееся разнообразие этих выборок. В этом случае варианса средних величин $s^2_{\bar{x}}$ связана с вариансой индивидуальных значений s^2 равенствами $s^2_{\bar{x}} = \frac{s^2}{n}$, где n – одинаковая численность выборок.

Среднее квадратическое или стандартное отклонение

Используется как более точный показатель для характеристики изменчивости.

Среднее квадратическое отклонение обозначается греческой буквой σ (сигма).

4. Среднее квадратическое отклонение

a) для несгруппированных данных

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



б) для сгруппированных данных

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 \cdot f}{\sum f}}$$

σ представляет собой среднее квадратическое отклонение вариант ряда от средней величины

Этот показатель указывает, насколько в среднем каждая варианта отклоняется от среднего арифметического. Эта величина именованная, т.е. выражается в тех же единицах, что и (кг, см, % и т.д.).

Чем больше величина σ , тем выше изменчивость.

Среднее квадратическое отклонение

имеет единицы измерения, а также может принимать положительные и отрицательные значения, поскольку получается в результате извлечения квадратного корня.

С помощью СКО можно утверждать, что i -тое значение признака в совокупности находится в пределах:

$$\bar{x} - \sigma \leq x_i \leq \bar{x} + \sigma$$

Относительные показатели вариации применяются для решения следующих задач:

- сравнение степени вариации различных вариационных рядов**
- характеристика степени однородности совокупности**

Коэффициент вариации

$$V = \frac{\sigma}{\bar{x}} \cdot 100\%$$

Характеризует долю усредненного значения отклонений от средней величины. При этом совокупность считается однородной, если V не превышает 33%

Относительный показатель

Коэффициент вариации (C_v) показывает, какой процент от средней арифметической составляет σ .

- С помощью коэффициента вариации можно сравнить изменчивость двух и более групп в отношении признаков, единицы измерения которых могут быть разными (например, число колосков в колосе, масса зерна, длина стебля и т. д.).
- При характеристике совокупности коэффициент вариации является дополнительным показателем и должен применяться с основными параметрами и σ .



Абсолютные показатели вариации являются именованными числами, т.е. имеют единицу измерения (такую же как и значения признака). Поэтому их нельзя использовать:

1) для сравнения степени вариации одного и того же признака в двух группах с разным уровнем средних;

2) для сравнения вариации двух различных признаков в одной группе.

Вся изменчивость признака лежит от среднего арифметического в пределах $\pm 3,3 \sigma$ ($\pm 3,3 \sigma$).

Это называется *правилом «плюс-минус трех сигм»*.

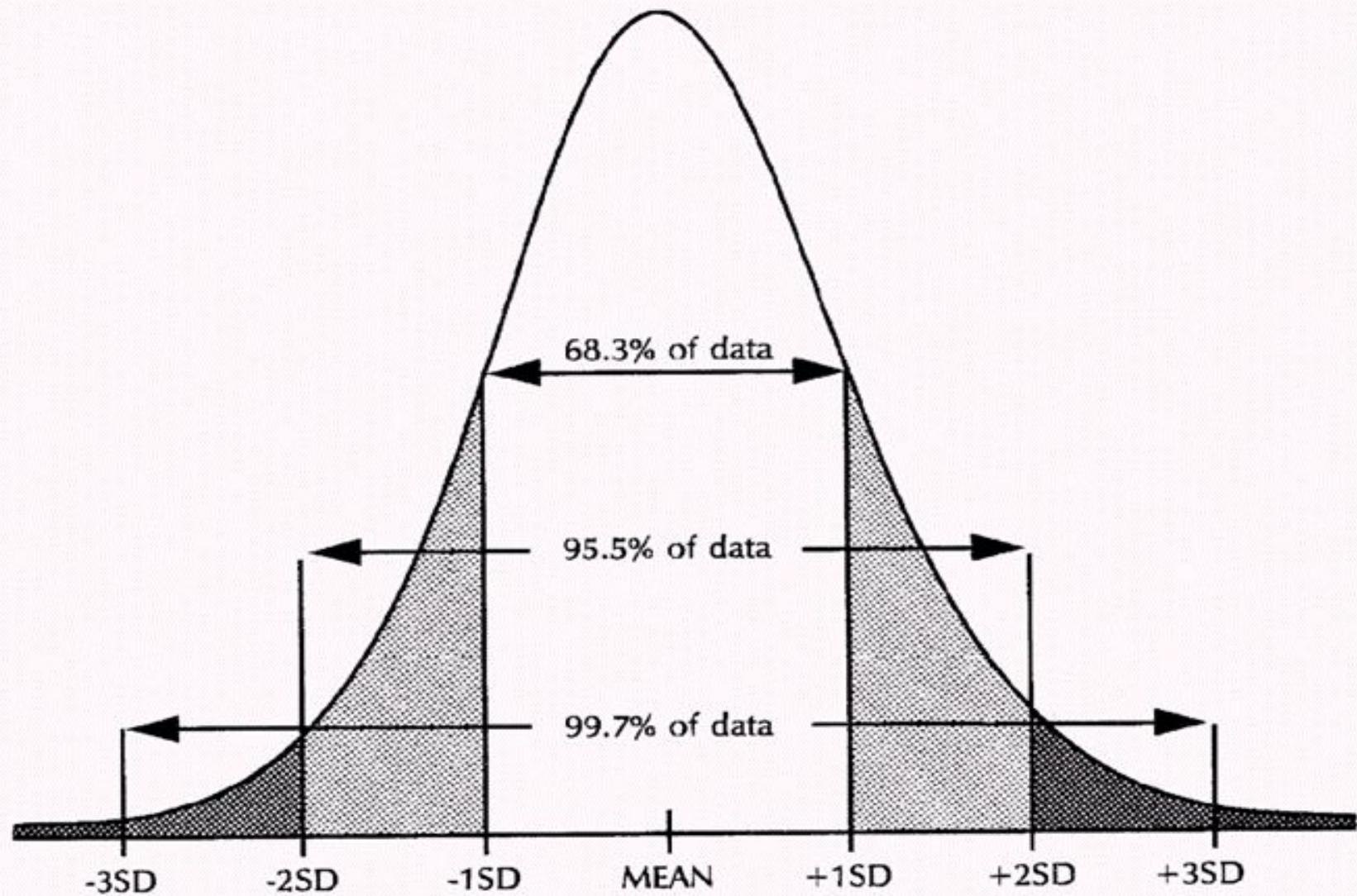
Поэтому средняя арифметическая, увеличенная и уменьшенная на три сигмы, дает практически крайние значения признака при нормальном распределении объектов в совокупности.

В условиях нормального распределения существует зависимость между величиной σ и количеством наблюдений:

в пределах $\bar{x} \pm \sigma$ располагается 68,3 % наблюдений;

в пределах $\bar{x} \pm 2\sigma$ располагается 94,5 % наблюдений;

в пределах $\bar{x} \pm 3\sigma$ располагается 99,7 % наблюдений.



Средняя

На практике почти не встречаются отклонения, которые превышают 3σ . Отклонение в 3σ может считаться максимальным

При помощи этого правила можно получить примерную оценку σ :

$$\sigma \approx \frac{x_{\max} - x_{\min}}{6}$$