

Основы научных исследований в садоводстве. Биометрия растений

Аниськина
Татьяна Сергеевна

Структура курса:

- *Раздел 1. Статистические параметры выборки*
- *Раздел 2. Анализ сопряженности варьирования признаков*
- *Раздел 3. Дисперсионный анализ:*
- *Раздел 4. Планирование экспериментов с садовыми растениями*

Научные исследования в любой отрасли деятельности основаны на теории планирования экспериментов и применении методов математической статистики для анализа экспериментальных данных.

- **Биометрия** - наука о применении математических методов для изучения биологических организмов.
- Предметом изучения биометрии являются математические методы, используемые для тех или иных суждений о биологических явлениях и процессах.



Задачи биометрии

- 1. Вычисление биометрических характеристик выборки.
- 2. Оценка достоверности выборочных биометрических характеристик: то есть, оценка степени их соответствия генеральным биометрическим характеристикам.
- 3. Оценка достоверности различий между выборками по тем или иным признакам.
- 4. Оценки степени влияния тех или иных факторов на признаки выборки.
- 5. Оценка степени сопряженности варьирования признаков.
- 6. Прогнозирование изменения тех или иных признаков в зависимости от изменения других признаков или факторов.

Статистические параметры выборки



Понятие о совокупности.

- **Совокупность** - всякое множество отдельных объектов, отличающихся друг от друга и в то же время сходных в некоторых существенных отношениях.
- **Генеральная совокупность** – теоретически бесконечно большая совокупность из всех единиц, которые могут быть к ней отнесены. (Например, все растения того или иного вида)
- **Выборочная** – совокупность конечного объема. Используется на практике.

Единица совокупности – x_i
Объем совокупности – n или N



- Единица совокупности характеризуется определенными **признаками**. Признаки варьируют (признак = переменная)
- Варьирующую величину, то есть, величину, изменяющуюся под влиянием многих случайных причин и принимающую разные значения, называют **случайной переменной x**
- Различие между единицами совокупности по тому или иному признаку (переменному) называется **вариацией** или **дисперсией** (рассеянием).
- **Варианта** - значение признака у той или иной единицы совокупности - x_i , где i – порядковый номер варианты.

Данные

Качественные

Категориальные
(неупорядоченные)

Порядковые
(упорядоченные)

Количественные

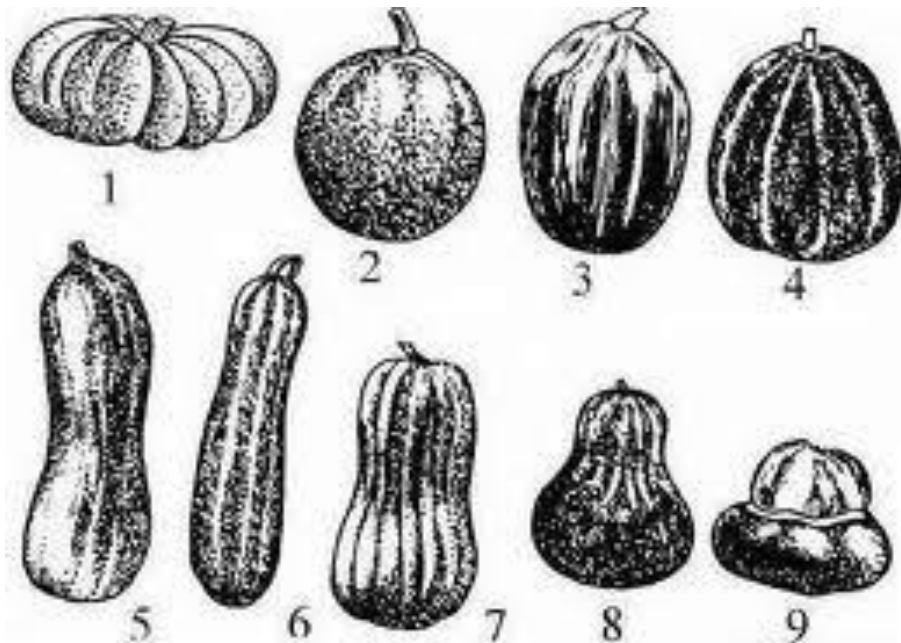
Дискретные (строго
фиксированные)

Непрерывные (зависят
от точности измерений)

Понятие о переменных (признаках)

- Типы вариации:

1) **качественная** вариация – признаки имеют очень ограниченный ряд состояний (форма, окраска);
Качественные признаки не возможно измерить.





2) *количественная* вариация

а) *дискретная*: различия между отдельными значениями случайной переменной выражаются целыми числами, без переходов (например, число растений, количество листьев, количество лепестков в цветке и т.п.); **5 шт.**

б) *непрерывная*: различия между отдельными значениями случайной переменной зависят от степени точности измерений (масштаба, интервала) количественного признака (например, масса плода, урожайность, диаметр венчика и т. п.). **0,8357 см**

Способы учета признаков – шкалы оценки

- **Шкала оценки** – это способ измерения состояния переменного.
- Существует **три типа** основных шкал оценки признаков: номинальная, порядковая и интервальная. Эти шкалы отличаются друг от друга по 2 основным свойствам:
 - 1) по наличию или отсутствию **правила ранжирования** состояний переменного;
 - 2) по наличию или отсутствию заданного **интервала** между состояниями переменного.

Номинальная (категориальная) шкала

- Нет правил ранжирования;
- Нет интервала между состояниями переменных;
- **Модальность** – состояние признака. (мономодальные – нет вариации; бимодальные – 2 модальности; полимодальные – 3 и более вариаций)
- Используют для оценки качественных признаков (состояния которого невозможно количественно измерить) и категоризированных (количественные признаки, искусственно превращенные в номинальные).
- **Исходными данными** для анализа являются частоты встречаемости признаков.

Достоинства и недостатки

- + универсальность (для качественных и количественных признаков)
- + простота и быстрота оценки
- + не нужны специальные приборы

- - низкая информативность
- - субъективность

Порядковая (ранговая) шкала.

- Есть правила ранжирования;
- Нет интервала между рангами;
- **Ранг R_i** – состояние признака, порядковый номер при ранжировании; В случае равенства x_j для нескольких объектов в выборке, рангом будет среднее арифметическое из соответствующих порядковых номеров этих переменных. Сумма всех рангов в выборке всегда должна быть равна сумме порядковых номеров.
- Чаще для качественных признаков (напр.: устойчивость к болезням – сильная, средняя...), которые можно упорядочить (ранжировать).

Достоинства и недостатки

- + относительная универсальность (для качественных и количественных признаков)
- + простота и быстрота оценки
- + относительная объективность

- - не высокая информативность
- - некоторая субъективность

Интервальная шкала

- Есть правила ранжирования;
- Есть интервал – есть начальная точка и единица измерения (гр, см, мм, Вт...)
- **Варианта x_i** – состояние переменной.
- Основная шкала для количественных признаков.

Достоинства и недостатки

- + объективность
- + наибольшая информативность
- - только для количественных признаков
- - необходимы инструменты

Унификация - сведение всех переменных к одной шкале

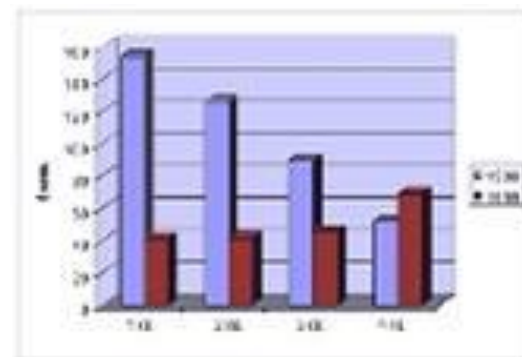
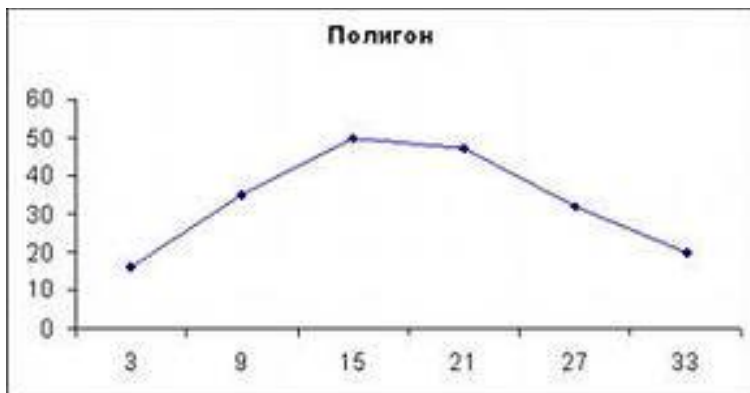
- **К номинальной шкале:**
 - + всегда осуществим;
 - потеря информации;
- **К порядковой шкале:**
 - + часто осуществим;
 - средняя потеря информации;
- **К интервальной:**
 - редко осуществим;
 - + наибольшее количество информации;



Вариационным рядом называют распределение частот встречаемости всех классов совокупности.

Изображается в виде вариационной кривой (кривой распределения):

- а) **полигон** частот распределения – на оси абсцисс наносятся середины классов; на оси ординат – частоты их встречаемости;
- б) **гистограмма** частот распределения признака – на Ох – границы классов, на Оу- частоты их встречаемости;



Статистические параметры делятся на 2 группы

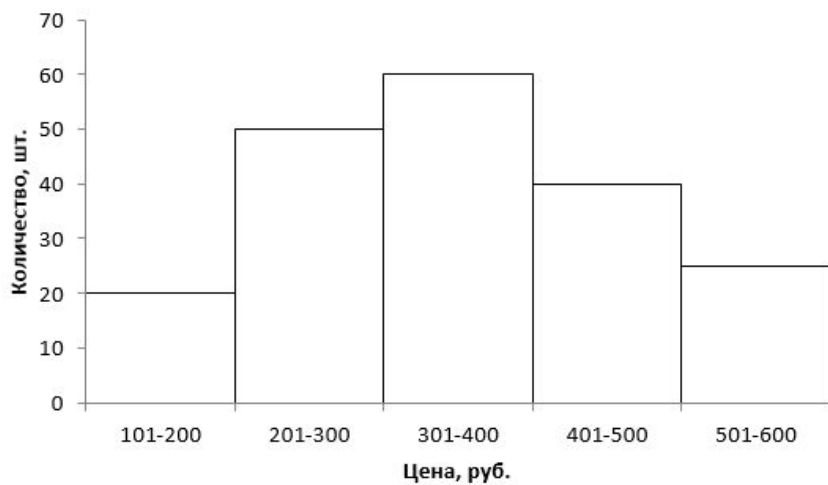
- Показатели *средней тенденции*:

- 1) Мода;
- 2) Медиана;
- 3) Среднее арифметическое.

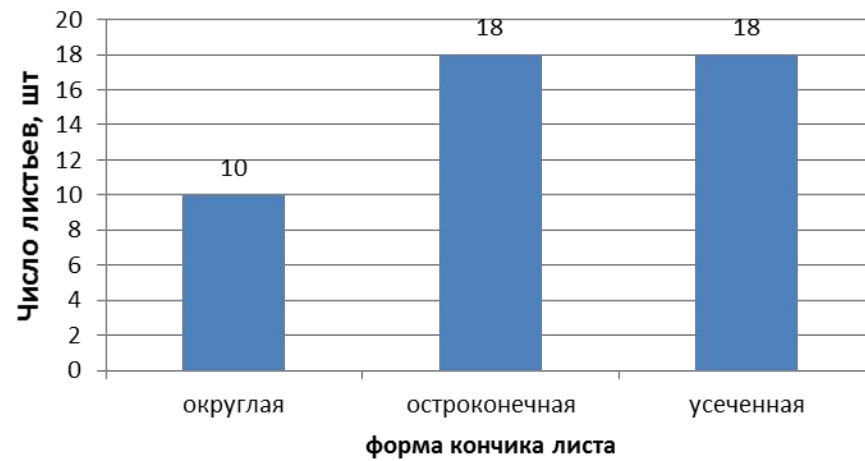
- Показатели *вариации*:

- 1) Размах вариации;
- 2) Среднее абсолютное отклонение;
- 3) Среднее квадратическое отклонение;
- 4) Дисперсия (варианса);
- 5) Коэффициент вариации.

Мода



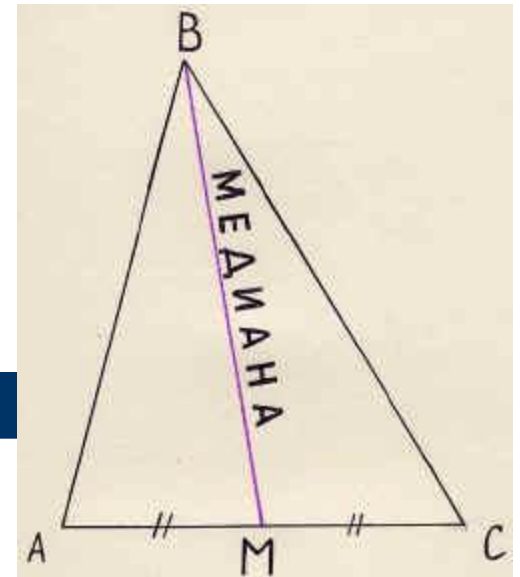
Распределение частот встречаемости листьев по форме кончика листовой пластинки



Мода

- 1) Мода – M_o – значение класса с максимальной частотой встречаемости;
- 2) Значение моды зависит от частот распределения классов.
- 3) Можно вычислить для любых признаков.
- 4) Количество мод в выборке – минимум = 1, максимум = числу классов.
- 5) Для количественных признаков M_o – среднее значение (середина) модального класса.

Медиана



Example 1:

(1) 5, (2) 8, ✓ 13, (1) 15, (2) 17

median = 13

Example 2:

✓ 5, ✓ 8, 15, 17

$$= \frac{8 + 15}{2} \quad \text{median} = 11.5$$

Медиана

- *Медиана* (Me) – значение варианты, которая находится точно в середине ранжированного вариационного ряда.
- Если объем выборки является четным числом, то медиана является средним значением двух соседних срединных вариантов.
- Если объем выборки является нечетным числом, то медиана является значением срединной (центральной) варианты.
- Свойства медианы: 1) медиана в выборке всегда одна; 2) медиана относительно устойчива, и наименее зависит от значений отдельных вариантов.
- Медиану можно вычислить только для признаков, оцененных в порядковой или в интервальной шкалах.

Среднее арифметическое

- Среднее арифметическое – частное от деления суммы всех вариантов выборки на объем выборки (N).
- Свойства средней арифметической:
 - 1) если каждую из вариантов совокупности увеличить или уменьшить на одну и ту же величину, то и средняя арифметическая соответственно уменьшится или увеличится на эту же величину;
 - 2) сумма разностей между отдельными вариантами и средней арифметической равна нулю;
 - 3) сумма квадратов отклонений вариант от средней арифметической всегда меньше суммы квадратов отклонений вариант от любой другой величины не равной средней арифметической.

- Особенности средней арифметической:
- 1) средняя арифметическая характеризует всю совокупность в целом, а не отдельные единицы совокупности;
- 2) средняя арифметическая имеет смысл только по отношению к качественно однородной совокупности;
- 3) средняя арифметическая характеризует только данную совокупность, экстраполировать её рискованно.
- 4) средняя арифметическая вычисляется только для признаков, измеренных в интервальной шкале.

Показатели вариации

- 1) *Размах вариации (изменчивости)* - (lim) – разница между максимальным и минимальным значениями признака в совокупности: $lim = X_{max} - X_{min}$;

Недостатки данного показателя:

- 1) очень не устойчивый (зависит только от крайних значений совокупности);
- 2) при равенстве размаха изменчивости двух выборок, распределение в них вариант может быть разным.

- *Среднее абсолютное отклонение* - (u) – это частное от деления суммы абсолютных отклонений всех вариантов от среднего арифметического на объем выборки N ;
- Недостаток: плохо отражает закономерности вариации признака в совокупности, поэтому им пользуются редко.

$$u = \frac{\sum |x_i - \bar{x}|}{N}$$

Дисперсия (варианса, σ^2) – это средний квадрат отклонений вариант от средней арифметической совокупности. Нет единицы измерения.

Число степеней свободы (df) – число независимых наблюдений

Для генеральной совокупности

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

Для выборочной совокупности

$$\sigma^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}{N - 1}$$

- *Среднее квадратическое отклонение* - корень квадратный из дисперсии:

$$\sigma = \sqrt{\sigma^2}$$

Измеряется в тех же единицах, что и признак. Среднее квадратическое отклонение может быть положительным и отрицательным числом ($\pm\sigma$).

- *Коэффициент вариации* (cv) - для сравнения вариации разных признаков. Это частное от деления среднего квадратического отклонения (σ) на среднюю арифметическую:

$$cv = \frac{\sigma}{\bar{x}} \times 100\%$$

Обычно выражается в процентах.

Основные статистические параметры выборки

- 1) объем выборки (N);
- 2) среднее арифметическое (\bar{x}) как наиболее важный показатель средней тенденции;
- 3) дисперсия (σ^2) как основной показатель вариации.