

Лекция 1

Тема:

***Математические идеи
в почвоведении***

ЦЕЛЬ КУРСА :

овладение современными методами статистического анализа данных, получаемых в результате почвенных и агрохимических исследований.

В процессе освоения курса студент должен овладеть умением применять результаты статистического анализа для интерпретации полученных данных и делать содержательные выводы.

МЕСТО КУРСА В ПРОФЕССИОНАЛЬНОЙ ПОДГОТОВКЕ ВЫПУСКНИКОВ

Программа курса «Математические методы в почвоведении» предусматривает цикл практических занятий, цель которых освоение алгоритмов решения различных статистических задач в применении к реальным почвенным данным. Освоение материала предусматривает постепенное нарастание сложности задач к концу практического курса, что позволяет в конце семестра приобрести достаточные навыки для самостоятельного применения различных статистических методов в научной работе студентами-почвововедами. Навыки, полученные на практических занятиях, закрепляются в результате выполнения самостоятельных работ, предусмотренных в программе курса.

Курс «Математические методы в почвоведении» проводится в 4 семестре подготовки специалистов, объем аудиторных часов – 36, из них лекции - 18 часа, практические занятия - 18 часа, **самостоятельная работа – 96 часов.** Общее количество часов –160.

ТРЕБОВАНИЯ К УРОВНЮ УСВОЕНИЯ СОДЕРЖАНИЯ КУРСА:

Студент должен свободно ориентироваться в выборе статистических методов при решении конкретных профессиональных задач. Навыки в овладении математическими методами в почвоведении студент получает на 12 практических занятиях и в результате выполнения 11 самостоятельных работ. Теоретическими знаниями студент овладевает на лекциях и в результате самостоятельной работы по программе курса с рекомендуемой литературой.

СПИСОК РЕКОМЕНДОВАННОЙ ЛИТЕРАТУРЫ

- Владимирский Б.М. Математические методы в биологии. Ростов-на-Дону: РГУ, 1983. -304 с.
- Владимирский Б.М. Горстко А.Б., Ерусалимский Я.М. Математика. Общий курс. СПб.: Изд-во «Лань», 2002. – 960 с.
- **Дмитриев Е.А. Математическая статистика в почвоведении. М.: Изд-во Моск. ун-та, 1972 - 292 с.**
- **Доспехов Б.А. Методика полевого опыта. М. Колос, 1968. -336 с.**
- Дюк В. Обработка данных на ПК в примерах. Санкт-Петербург: Питер, 1997. -240 с.
- Дюран Б., Одел П. Кластерный анализ. М.: Статистика. 1997. - 128 с.
- **Лакин Г.Ф. Биометрия. М.: высшая школа, 1990. – 352 с.**
- Лебедева Н.В., Дроздов Н.Н., Криволуцкий Д.А. Биоразнообразие и его оценка. М.: МГУ, 1999. – 99 с.
- **Лебедева Н.В., Рыбьянец Т.В. Математические методы в почвоведении. Задачи с методическими указаниями. Ростов-на-Дону: Биос, 2002. – 66 с.**
- Лоули Д. Максвелл Л. Факториальный анализ как статистический метод. М.: Мир. 1967. - 144 с.
- Применение методов математической статистики в почвоведении, мелиорации и сельском хозяйстве. Москва-Новочеркасск. – 1980. – 57 с.
- Рожков В.А. Почвенная информатика. М.: Агропромиздат, 1989. – 221 с.
- Рокицкий П.Ф. Биологическая статистика. Минск: Вышэйш. Школа, 1967. 328 с.
- Справочник по прикладной статистике/ Под ред. Э.Ллойда, У. Ледермана, Ю.Н. Тюрина. Т. 1,2. М.: Финансы и статистика. 1989, 1990. 510 с., 526 с.

План лекции:

1. Развитие математических идей в биологии и почвоведении.
2. Понятие генеральной совокупности и случайной выборки. Признаки. Выборочные распределения.
3. Графическое представление экспериментальных данных.

Развитие математических идей в биологии и почвоведении

Применение математического аппарата в исследованиях натуралистов и естествоиспытателей в XVII – XIX вв. Эволюционная теория и теория вероятности. Развитие математических идей в антропологии, генетике, физиологии, популяционной биологии, почвоведении и агрохимии. Появление раздела биологии – биометрии. Работы Ф. Гальтона и К. Пирсона. Роль англо-американской школы в разработке математического аппарата биологии. Работы В. Госсета, Р. Фишера. Роль Ю.А. Филипченко, В.И. Романовского, А.Н. Колмогорова, С.С. Четверикова, Л.О. Карпачевского, Е.А. Дмитриева, Б. А. Доспехова и др. в становлении традиций преподавания и применения математических методов в почвенных и биологических исследованиях в нашей стране. Решение сложных математических задач и моделирование в почвоведении и биологии в связи с развитием вычислительной техники.

В глубокой древности, с образованием государств возникла необходимость в учете населения, численности войск, имущества граждан, земельных наделов, поголовья скота и т.д..

Такие сведения еще за несколько тысячелетий до н.э. собирались в Китае, Египте, Древнем Риме. В Средние века эпизодически проводились переписи населения, домашнего имущества, земельных угодий; велся учет лиц, пригодных к военной службе.

С развитием торговли и международных товаро-денежных отношений все более расширялся круг учитываемых явлений и собираемых о них сведений. Стали собираться данные о ввозимой продукции, ценах на товары и т.д.

XVII в

Теория
вероятности

П. Ферма (1601-1665),
Б. Паскаль (1623-1662)
Х. Гюйгенс (1629-1695)
К. Гаусс (1777-1855),
П. Лаплас (1749-1827)

Математическа
я статистика

В. Петти (1623-1687),
Г. Галилей (1564-1642)
Борелли (1608-1679)
А. Кетле (1796-1874) и др.

Адольф Кетле

(22.02.1796 – 17.12.1874)

Формирование статистики в научную дисциплину связано с деятельностью бельгийского ученого впервые применившего научные методы обработки статистических данных.

Одним из первых показал, что случайности наблюдаемые в живой природе вследствие их повторяемости, обнаруживают тенденцию, следуя определенной закономерности, некоторую можно исследовать и описать точным и математическими методами.

Труды:

1835 «О человеке и развитии его способностей, или Опыт социальной физики»

1848 «О социальной системе и законах, управляющих ею»

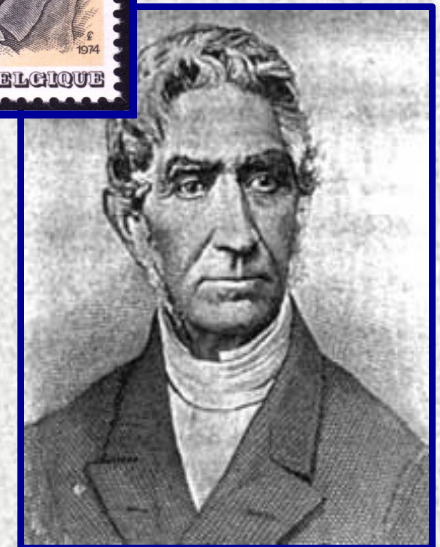
1871 «Антропология»



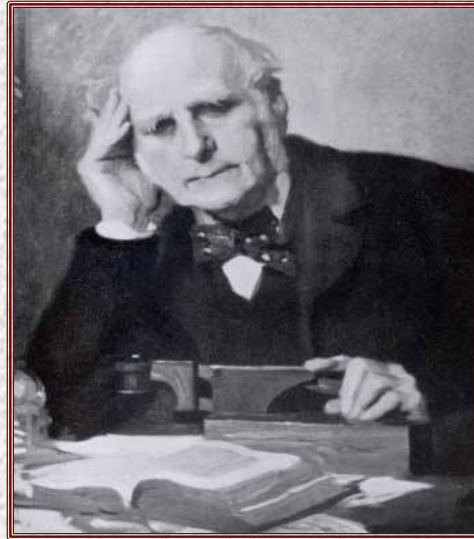
Совместно с ведущими математиками своего времени А. Кетле организовал так называемую Центральную статистическую комиссию, пропагандировавшую важность проведения серьезных статистических исследований.



С именем Кетле в истории науки связан переход социальной статистики от сбора и количественного описания данных к установлению постоянных корреляций показателей, или статистических закономерностей. Выявленные им с помощью математического вероятностного анализа постоянные соотношения показателей Кетле трактовал как объективные социальные законы.



Ф. Гальтон (1822-1911)



**Английский
ученый
основатель
школы
биометриков.**

**Создал
основной
математич
еский
аппарат
биометрии.**

**Разработал
методику
регрессионного
и
корреляционног
о анализа.**

**Впервые применил
статистический
метод Кетле к
проблемам
наследственности
и изменчивости.**



**КАРЛ ПИРСОН
(1857-1936)**

Пирсон считается одним из основателей статистики. В молодости у него появился интерес к проблемам наследственности, евгеники, общим вопросам биологии и возможности применения методов статистики для их изучения. Значительную часть своих усилий употребил на разработку и применение статистических методов в биологии. Он считается одним из отцов современной статистики.

Его интерес сформировался под влиянием эволюционной теории Дарвина, которую он хотел проверить с помощью различных методов.

Вместе с такими философами, как Юм и Мах, Пирсон разделял взгляды на причинность как сопряженную вариацию ("корреляция").

Ободренный сэром Фрэнсисом Гальтоном и Уолтером Велдоном, он предложил математическую формулировку идеи корреляции, о которой размышлял и сам Гальтон. Результатом явился широко известный коэффициент корреляции Пирсона.

С именем Пирсона связано понятие среднего квадратического отклонения, обоснование метода хи-квадрат. Он развил теорию линейной и нелинейной корреляции, разработал методiku построения множественной регрессии.

Совместно с Гальтоном и Уэльдоном основал в 1901 г выпуск журнала «Биометрика».





В. Госсет - «Стьюдент» (1876-1937)

Открыл закон распределения выборочных средних в зависимости от объема выборки. Обоснована теория «малой выборки».

Описанный закон распределения оказался применим к выборкам любого размера.



Критерий Стьюдента направлен на оценку различий величин средних значений двух выборок, которые распределены по нормальному закону. Одним из главных достоинств критерия является широта его применения. Он может быть использован для сопоставления средних у связных и несвязных выборок, причем выборки могут быть не равны по величине.

t-критерий был разработан Уильямом Госсетом (1876-1937) для оценки качества пива на пивоваренных заводах Гиннеса в Дублине (Ирландия).

В связи с обязательствами перед компанией по неразглашению коммерческой тайны (руководство Гиннеса считало таковой использование статистического аппарата в своей работе), статья Госсета вышла в 1908 году в журнале «Биометрика» под псевдонимом «Student» (Студент).

Р.Э. Фишер (1890-1962)



Работал научным сотрудником Ротамстедской сельскохозяйственной опытной станции, а с 1933 г профессором кафедры прикладной математики Лондонского университета, с 1943 по 1957 заведовал кафедрой генетики в Кембриже.

Разработал метод комплексной оценки действия факторов и их возможных сочетаний на резульативный признак – метод основанный на разложении дисперсии, получивший название дисперсионного анализа.

Ввел науку такие понятия как: «варiances», «число степеней свободы», «даты».

Заложил основы теории планирования эксперимента.

Первые опыты по использованию статистических методов в почвоведении относятся к 20-м гг. текущего столетия (Чириков, Малюгин, 1926; Качинский, 1926, 1927; Астапов, 1928; Соколов, 1929; Изюмов, 1930 и др.).

Этому немало способствовало появление литературы по математической статистике прикладного характера (Голубев, 1920; Сапегин, 1922; Филиппченко, 1926; Поморский, 1927 и др.).

***До конца 50-х начала 60-х гг.
статистические методы
использовались очень скромно, в этот
период появляется ряд работ весьма
ярких и не потерявших своего значения
до настоящего времени (Сердобольский,
1937, 1952; Филиппова, Сердобольский,
1937; Важенин и др., 1959, 1961; и др.).***

Интенсивное внедрение статистических методов в почвоведение, было инициировано и подготовлено работами по прикладной статистике в смежных областях науки, в том числе сельскохозяйственной

(Немчинов, 1945; Перегудов и др., 1948; Федоров, 1957; Финни, 1957; Фишер, 1958; Снедекор, 1961; и др.).

- ✓ Понятие генеральной совокупности и случайной выборки.
- ✓ Признаки: качественные и количественные (непрерывные и дискретные).
- ✓ Выборочные распределения.

Испытание осуществление какого-нибудь определенного комплекса условий, который может быть воспроизведен сколь угодно большое число раз.

Под **комплексом условий** понимают определенную методику выделения объекта изучения, способ его опробования и получение конечного результата измерений.

- Величины, в процессе испытаний не меняющие своих значений, называются постоянными.
- Если величина может менять по крайней мере два различных значения, то она называется переменной.

- Если в некотором интервале переменная может принимать любые значения (содержит все действительные числа), она называется *непрерывной*.
- *Дискретная* переменная может принимать лишь некоторые значения, разделенные промежутками.

Явления, происходящие в результате испытаний, в математической статистике и теории вероятности называются *случайным событием*.

Переменная величина, значение которой до проведения испытаний заранее не может быть точно предсказано называется *случайной величиной*.

Совокупность из которой отбирается некоторая часть ее членов для совместного изучения называется **генеральной.**

N – объем генеральной совокупности:

$$N \rightarrow \infty$$

Часть генеральной совокупности отобранная тем или иным способом называется **выборочной совокупностью (или выборкой).**

n – объем выборки:

$$**$n \geq 2$**$$

Выборочный метод

является основным при изучении статистических совокупностей.

Преимущества:

- сокращает время и затраты труда,
- позволяет получать информацию о таких совокупностях сплошное обследование которых практически невозможно или не целесообразно.

Выборка должна быть представительной – **репрезентативной** (от лат. represento – представляю), т.е. возможно полнее отображать структуру генеральной совокупности.

Рендомизация (от англ. random - случай) – случайный отбор вариант из генеральной совокупности, что обеспечивает равную возможность для всех членов генеральной совокупности попасть в состав выборки.

Способы отбора вариант из генеральной совокупности:



- Повторный
 - *возвращение учтенных единиц в генеральную совокупность*
- Бесповторный
 - *учтенные единицы в генеральную совокупность не возвращаются*

Виды отбора единиц из генеральной совокупности:

- Типический (или групповой);
- Серийный (или гнездовой);
- Механический.

Шкалы измерений:

- Именная (или классификационная);
- Порядковая шкала;
- Интервальная шкала;
- Шкала отношений.

Классификация признаков:

- Качественные
 - Признаки, поддающиеся измерению на именных шкалах
- Количественные
 - Признаки, которые могут быть ранжированы по какому-либо принципу
- Порядковые
 - Признаки, поддающиеся количественной оценке

Если в основе классификации лежит группировка данных, то признаки делят на:

- Альтернативные;
- Порядковые;
- Ранговые.

Общей основой классификации признаков являются мера и счет, в зависимости от этого признаки бывают:

- Признаки
 - Счетные или меристические
 - мерные
 - или метрические
 - Непрерывные
 - (могут принимать любые числовые значения в пределах: от-до)
 - Дискретные
 - (их значения всегда выражаются только целыми числами)

Процесс систематизации или упорядочения, первичных биометрических данных в целях извлечения заключенной в них информации, обнаружения закономерности, которой следует изучаемое явление или процесс называется группировкой. Она может быть простой и сложной.

Группировка по одному признаку называется простой (простая таблица), по нескольким признакам - сложной (корреляционная таблица).

Статистические ряды:

- Атрибутивные
- Вариационные
- Динамики
- Временные

Вариационным рядом называется ряд чисел, показывающий закономерность распределения единиц изучаемой совокупности по ранжированным (франц. Ranger – выстраиваться в ряд по ранжиру, т.е. по росту) значениям варьирующего признака.

Двойной ряд чисел, показывающий, каким образом числовые значения признака (x_i) связаны с их повторяемостью (p_i) в данной совокупности, называется вариационным рядом или рядом распределения.

Числа, показывающие, сколько раз отдельные варианты встречаются в данной совокупности называются **частотами или весами вариант (p или f)**. (Выражаются целыми числами или в %).

$$\Sigma p = n, \text{ где:}$$

Σ - **суммирование частот вариационного ряда;**
 n – объем выборочной совокупности.

Графическое представление

экспериментальных данных:

гистограммы, полигоны, кривые

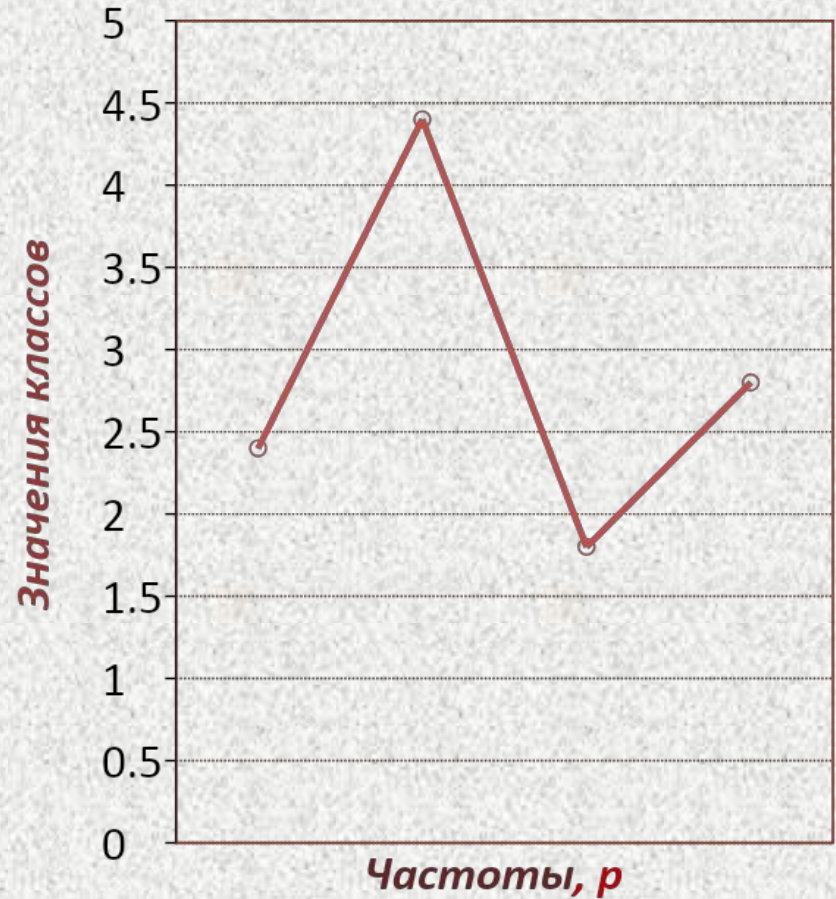
накопленных частот, ящички с усами,

круговая диаграмма и др.

Кривая распределения

Линейный график, получающийся при соединении прямыми линиями геометрических точек, связывающих значения классов с их частотами называется

вариационной кривой
(или кривая распределения).



Полигон распределения частот

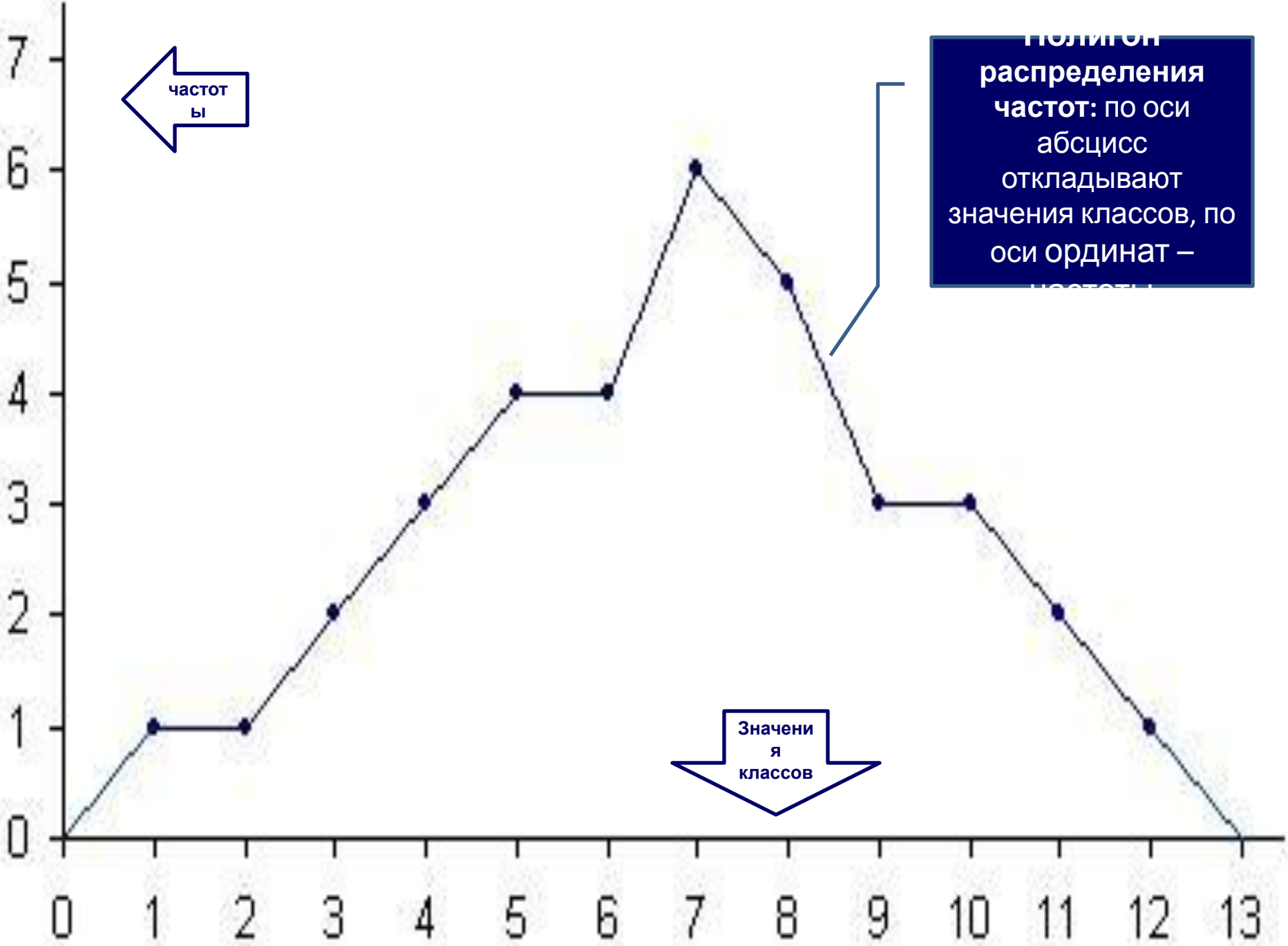
При построении графика безинтервального вариационного ряда по оси абсцисс откладывают значения классов, по оси ординат – частоты.

Высота перпендикуляров, восстановленных по оси абсцисс соответствует частотам классов. Соединяя вершины перпендикулярными линиями получаем полигон распределения частот.

← частоты

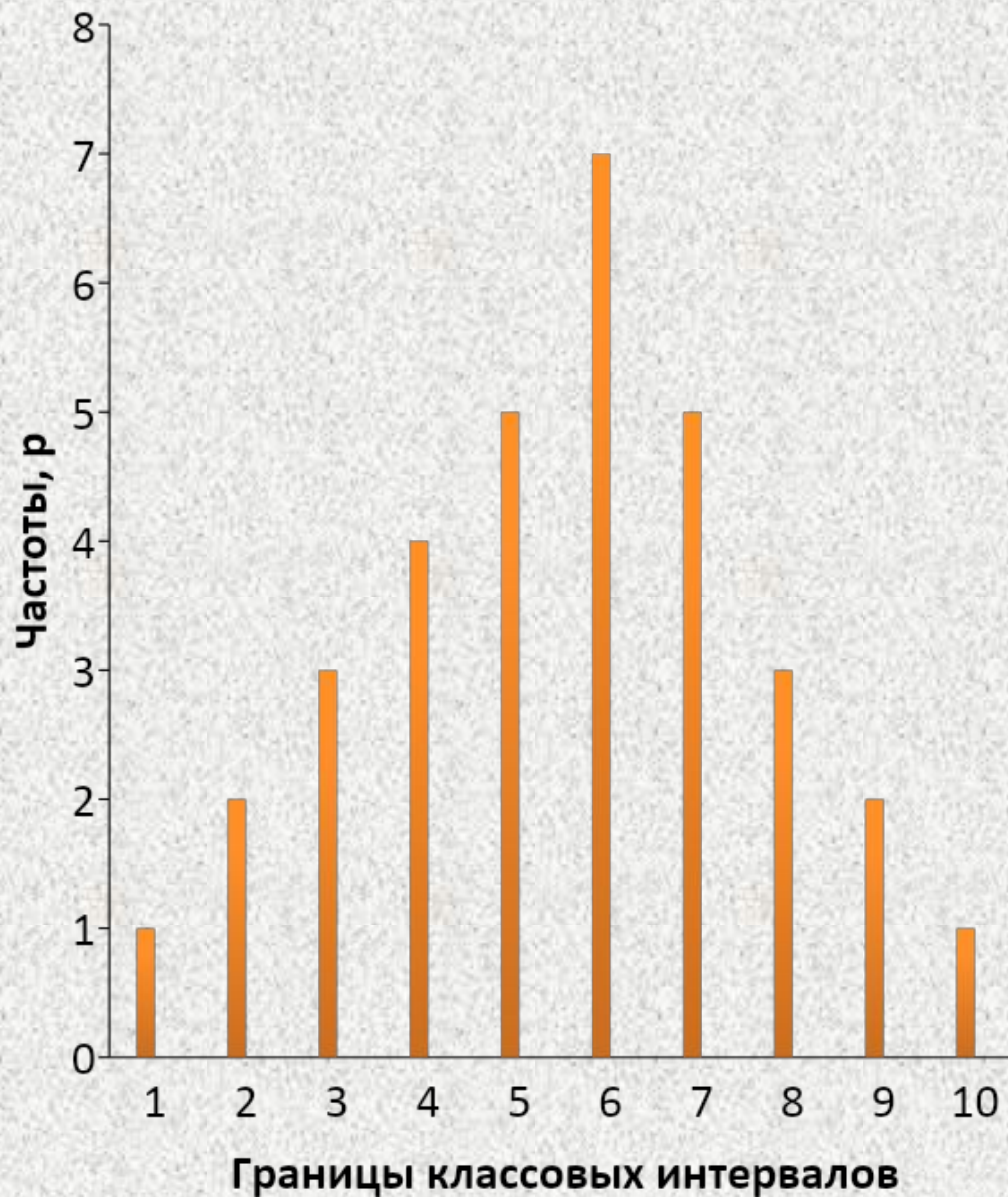
Полигон
распределения
частот: по оси
абсцисс
откладывают
значения классов, по
оси ординат –
частоты

Значения
классов
↓



Гистограмма распределения частот

При построении графика интервального вариационного ряда по оси абсцисс откладывают границы классовых интервалов, по оси ординат – частоты.



Эмпирический ряд распределения можно изобразить в виде кумуляты или огивы.

Накопленные частоты получаются последовательным суммированием или кумуляцией (от лат *cumulo* - накапливаю) частот в направлении от минимальной классовой варианты или от 1 класса до конца вариационного ряда.

Кумулята получается, если по оси абсцисс откладывают значения классов, а по оси ординат накопленные частоты.

Огива. Если накопленные частоты откладывают по оси абсцисс, а значения классов по оси ординат.

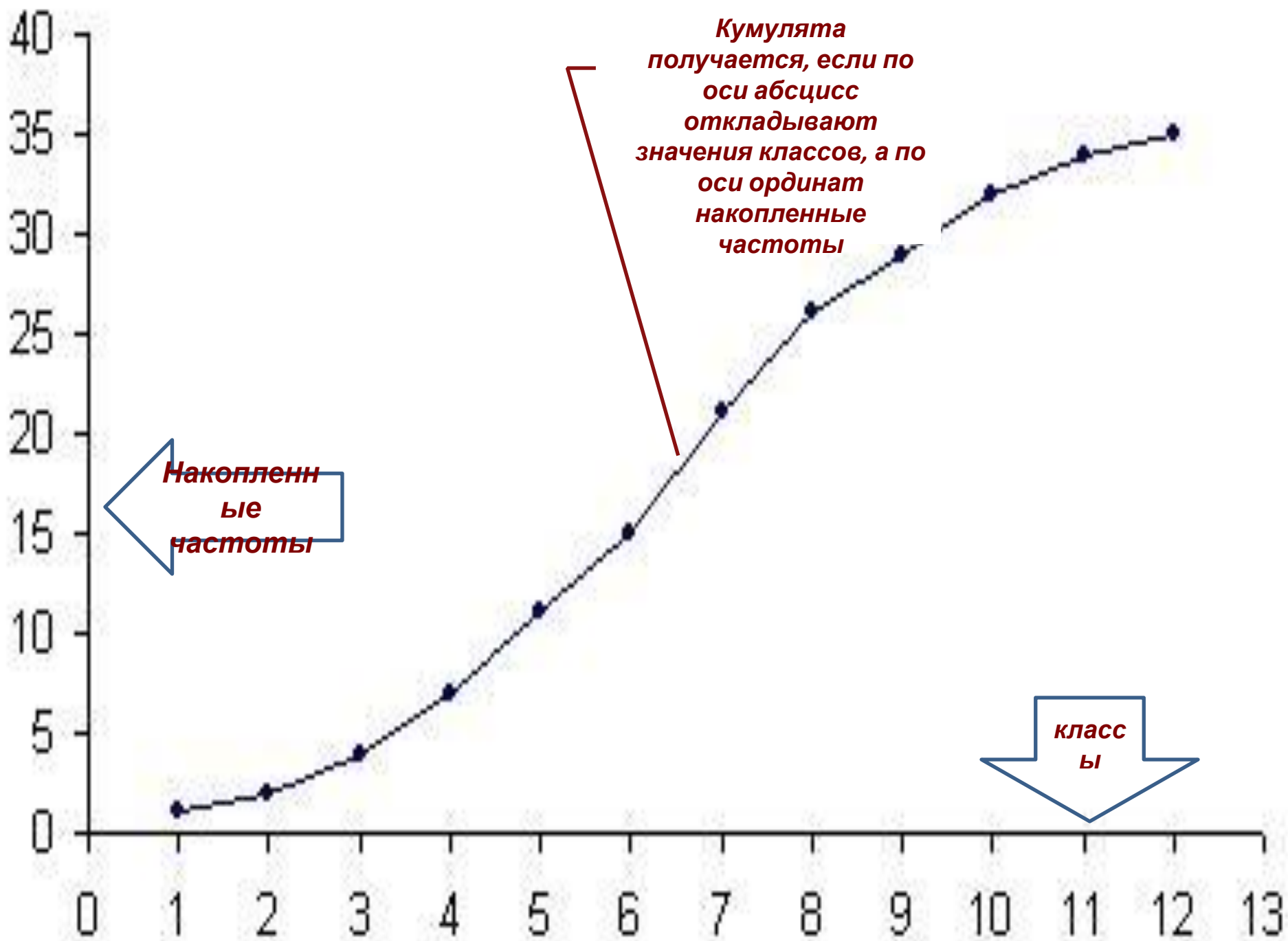


Диаграмма «Ящик с усами»

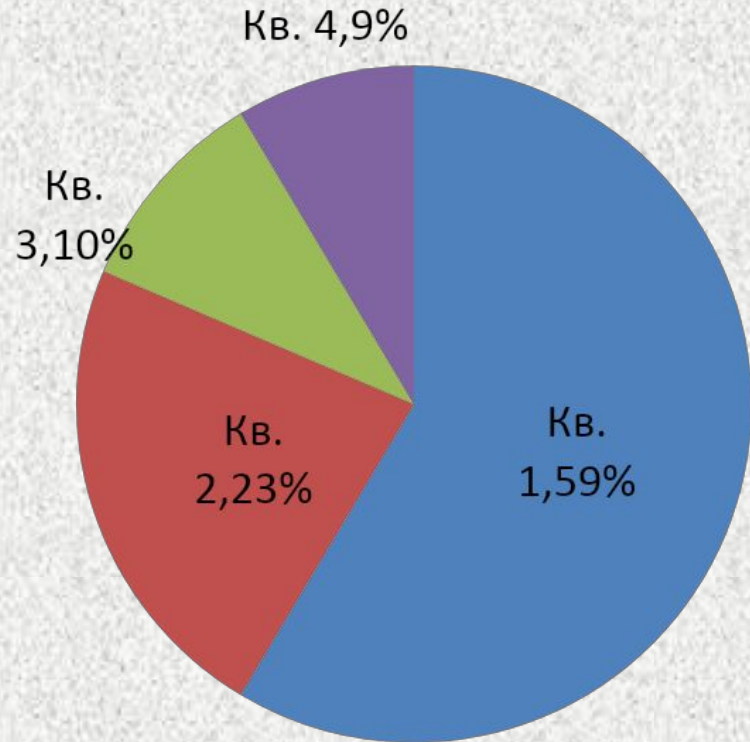
Результаты выборочных наблюдений могут быть представлены в виде эмпирически представленных квантилей.

На диаграммах размаха этого типа вокруг средней точки (т.е. среднего или медианы) рисуются прямоугольник, представляющий выбранный диапазон (т.е. стандартную ошибку, стандартное отклонение, минимум-максимум или константу), и отрезок, также отражающий выбранный диапазон, концы которого расположены вне прямоугольника (см. следующий рисунок).

Круговая диаграмма

Термин "*круговые диаграммы*" впервые был использован Хаскеллом в 1922 году.

На этих графиках пропорции отдельных значений переменной X представлены в виде круговых секторов.



Стьюде́нт из книги: Ермолаев О. Ю. «Математическая статистика для психологов» - 4-е издание испр. – М.: Московский психолого-социальный институт: Флинта, 2006 г.

Благодарю за внимание!