




Лекция по медицинской статистике

План лекции

- 1. Значение статистики для медицины и здравоохранения
- 2. Этапы статистического исследования
- 3. Описательная статистика (средние величины)
- 4. Сравнительная статистика




**Статистика – наука,
изучающая количественные
закономерности**

**материальных
явлений в неразрывной
связи с их качественной
стороной.**




Статистика:

- это инструмент для анализа экспериментальных данных и результатов популяционных исследований;
- это язык, с помощью которого исследователь сообщает полученные им результаты и благодаря которому он понимает медико-статистическую информацию;
- это элемент доказательной медицины;
- это база для обоснования принятия управленческих решений.




**Медицинская
статистика - раздел**

**статистики, изучающий
состояние здоровья
населения и общественное
здравоохранение**



Статистика
медико-
биологических
исследований



**ПРИМЕНЕНИЕ
СТАТИСТИЧЕСКИХ МЕТОДОВ
В МЕДИЦИНЕ В
ЗАВИСИМОСТИ ОТ УРОВНЯ
РЕШАЕМЫХ ЗАДАЧ**

Трактовка нормы и патологии

Популяционный уровень

Оценка состояния здоровья популяции (смертность, заболеваемость и т. д.)

Шкала оценки уровня смертности (коэффициент смертности на 1000 жителей).

До 10 - низкий

10-14,9 - средний

15-24,9 - высокий

25-34,9 - очень высокий

35 и более - чрезвычайно высокий

Индивидуальный уровень

Оценка состояния здоровья пациент (клинико-лабораторные показатели)

Нейтрофилы:

Число клеток в тыс. в 1 мкл крови

палочкоядерные

40—300

сегментоядерные

2000—5500

Диагностика заболеваний

Популяционный уровень

Разделение населения по группам здоровья

I — здоровые с нормальным уровнем функций; II — здоровые, но имеющие функциональные отклонения, а также сниженную сопротивляемость к острым и хроническим заболеваниям; III — больные хроническими болезнями в состоянии компенсации; IV — больные хроническими болезнями в состоянии субкомпенсации; V — больные хроническими болезнями в состоянии декомпенсации.

Индивидуальный уровень

Дифференциальная диагностика

Симптомы аллергического ринита	Частота симптомов $P \pm m$, %
Выделения из носа	96,2±3,8
Постназальный синдром	76,9±8,4
Чихание	96,2±3,8
Затруднение носового дыхания	96,2±3,8
Слезотечение	73,1±8,9
Зуд в глазах	80,8±7,9
Покраснение глаз	80,8±7,9
Першение в горле	88,5±6,4
Нарушение сна	88,5±6,4
Нарушение дневной активности ИИС	100,0±0,01

Прогнозирование процессов

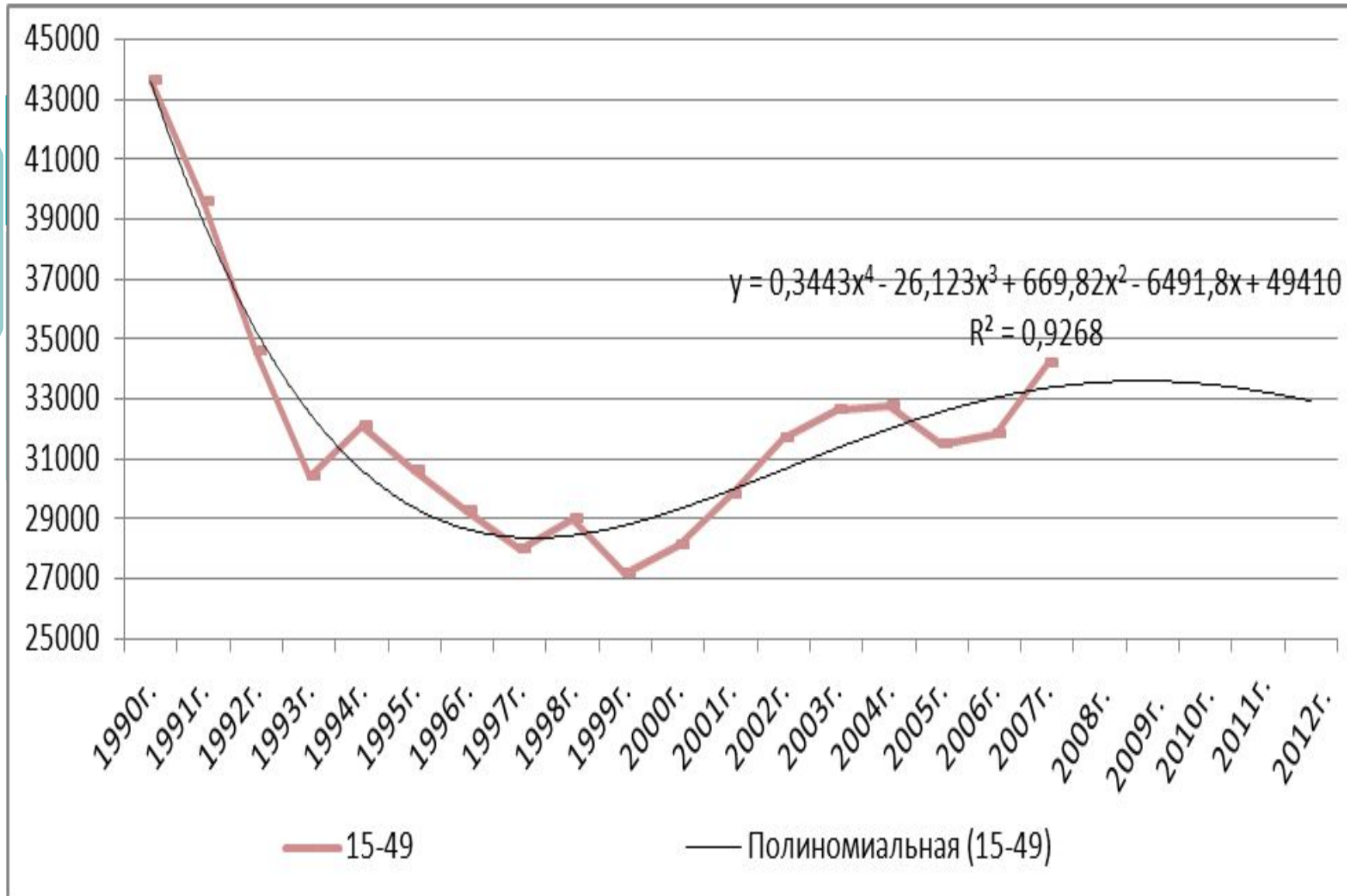
**Популяционный
уровень**

Отдаленные
результаты целевых
программ

**Индивидуальный
уровень**

Исход заболевания
отдельного больного

Динамика числа живорождений с прогнозом до 2012 г.



Выбор подходящего воздействия

**Популяционный
уровень**

Разработка целевых программ

**Индивидуальный
уровень**

Подбор индивидуальной тактики лечения


Организация медицинской помощи

Популяционный уровень

Анализ деятельности системы здравоохранения и общественного здоровья

Индивидуальный уровень

Анализ результатов своей работы




**Планирование и проведение
медицинских исследований
Подготовка публикаций и
сообщений**

**! Чтение и понимание медицинских
сообщений**

Анализ результатов исследования.

1. Описательная статистика (средние величины, относительные величины, параметры разброса)
2. Сравнительная статистика (выбор критерия для сравнения)
3. Определение связей между признаками (корреляционный анализ)
4. Прогнозирование (регрессионный, дискриминантный анализ, оценка выживаемости)
5. Классификация (кластерный, факторный анализ)



Этапы статистического исследования.

Этапы статистического исследования:

I. Формирование цели и задач исследования.

II. Организация исследования.

III. Сбор информации.

IV. Обработка информации.

V. Анализ результатов исследования.

VI. Внедрение результатов исследования в практику и оценка эффективности внедрения.

Формирование цели и задач исследования

Цель - отвечает на вопрос зачем проводится данное исследование.

Задачи исследования - дают ответ на вопрос как будет достигнута цель.

Организация исследования:

План исследования

предусматривает методику проведения исследования, дает раскладку организационных вопросов (что, где, когда, сколько?).

Определяет субъектов исследования.

Организация исследования:

Программа исследования (отвечает на вопрос: **как** делать?) состоит из трех главных компонентов :

- 1) программы сбора материала;
- 2) программы его разработки (табличной сводки);
- 3) программы анализа.

Прежде всего устанавливается **объект исследования и единица наблюдения.**

Организация исследования:

Под **объектом наблюдения** понимают статистическую совокупность, состоящую из отдельных предметов или явлений - единиц наблюдений, взятых в определённых границах времени и пространства.

Единица наблюдения - первичный элемент статистической совокупности, являющейся носителем признаков, подлежащих регистрации, изучению в ходе исследования.

Учетные признаки – признаки подлежащие регистрации в ходе статистического исследования.

Учетные признаки

```
graph TD; A[Учетные признаки] --> B[Качественные]; A --> C[Количественные]; B --> D["Альтернативная (номинальная) шкала (пол)"]; B --> E["Шкала рангов (порядковая) (стадии болезни)"]; C --> F["Интервальные (шкала Цельсия)"]; C --> G["Относительные шкалы (наличие нулевой точки)"];
```

Качественные

Количественные

Альтернативная
(номинальная)
шкала (пол)

Шкала рангов
(порядковая)
(стадии болезни)

Интервальные
(шкала Цельсия)

Относительные
шкалы (наличие
нулевой точки)



Учетные
признаки



Факторные



Результативные

Способы
наблюдения

Непосредственное
наблюдение

Выкопировка
данных

Опрос


По времени
наблюдения

Текущее

Единовременное

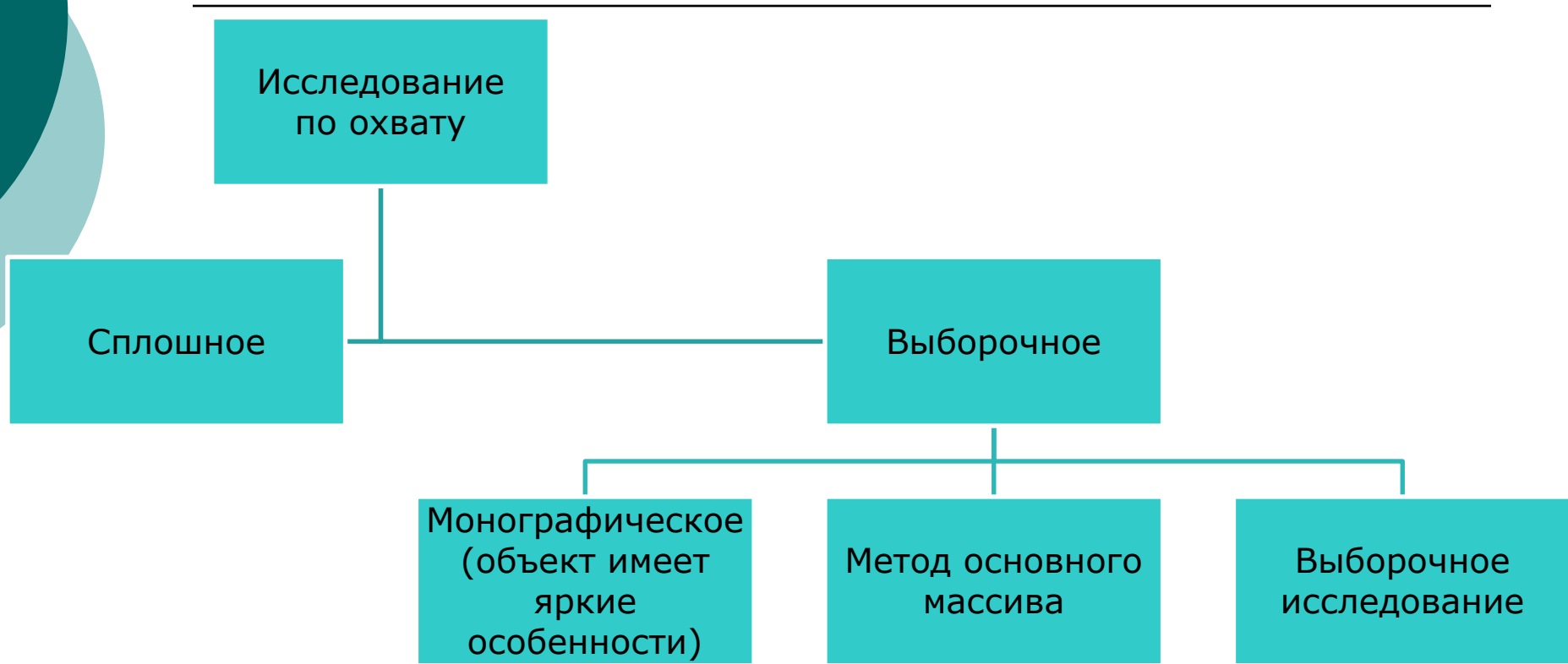
Текущее (непрерывное) наблюдение

- предусматривает регистрацию данных по мере их возникновения за какой-либо длительный промежуток времени.
- Например: регистрация актов гражданского состояния ЗАГСами за год



Единовременное (прерывное) наблюдение

- предусматривает регистрацию данных в один момент времени, или по состоянию на один момент времени, так называемый критический момент наблюдения.
- Таким образом, проводится сбор данных при переписях населения.





Монографический метод


- применяется для подробного описания объекта, имеющего какие-либо яркие особенности.
- Например, медико-социальное обследование национальностей Крайнего Севера

Метод основного массива

- предусматривает обследование контингентов, которые могут быть сосредоточены на конкретном объекте
- Например: изучение госпитализированной заболеваемости в конкретном стационаре.

Выборочное исследование

- Собственно выборочное исследование охватывает выборочную совокупность или просто выборку из генеральной совокупности.



```
graph TD; A[Репрезентативность выборки] --- B[Количественная]; A --- C[Качественная]
```

Репрезентативность
выборки

Количественная

Качественная



Средние величины

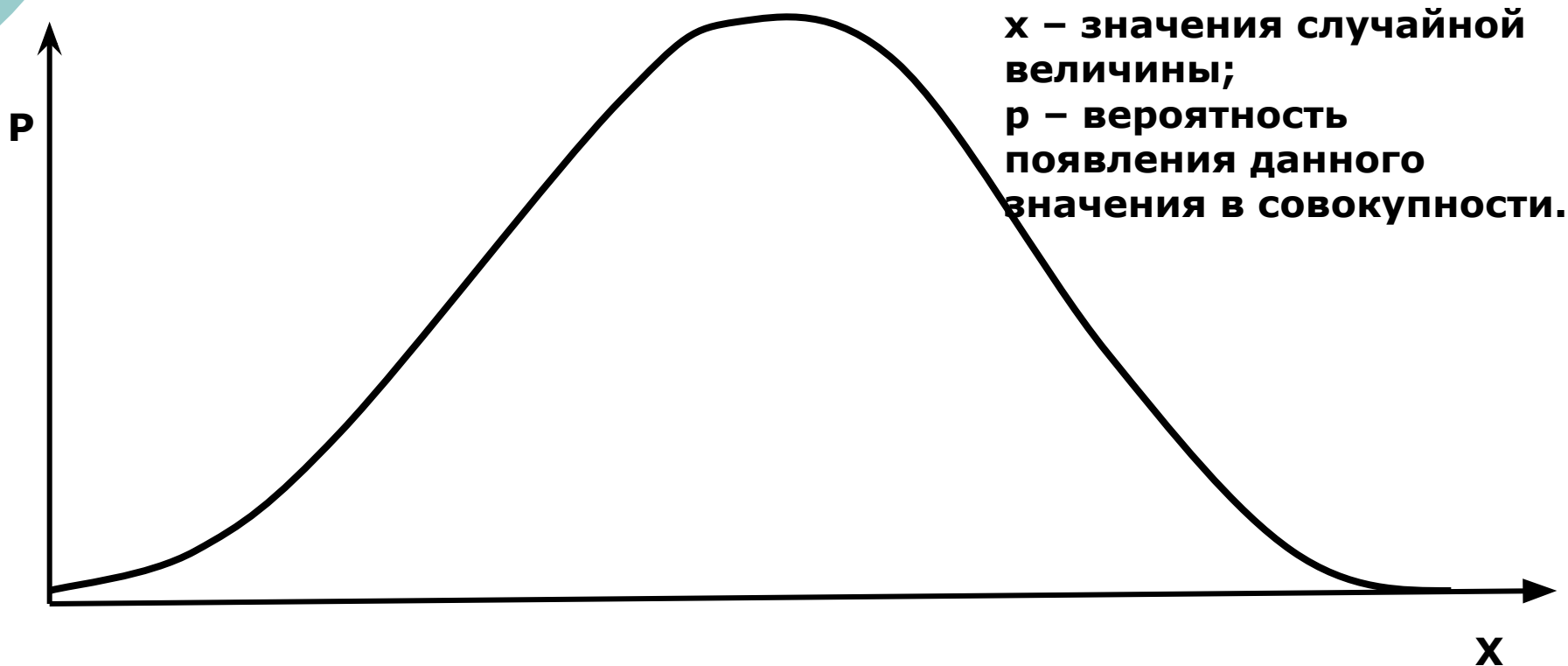
Вариационный ряд (frequency table)- ранжированный ряд распределения по величине какого-либо признака. Этот признак носит название варьирующего, а его отдельные числовые значения называются вариантами и обозначаются через "v". Число, показывающее, сколько раз данная варианта встречается в вариационном ряду, называется частотой и обозначается через "p"

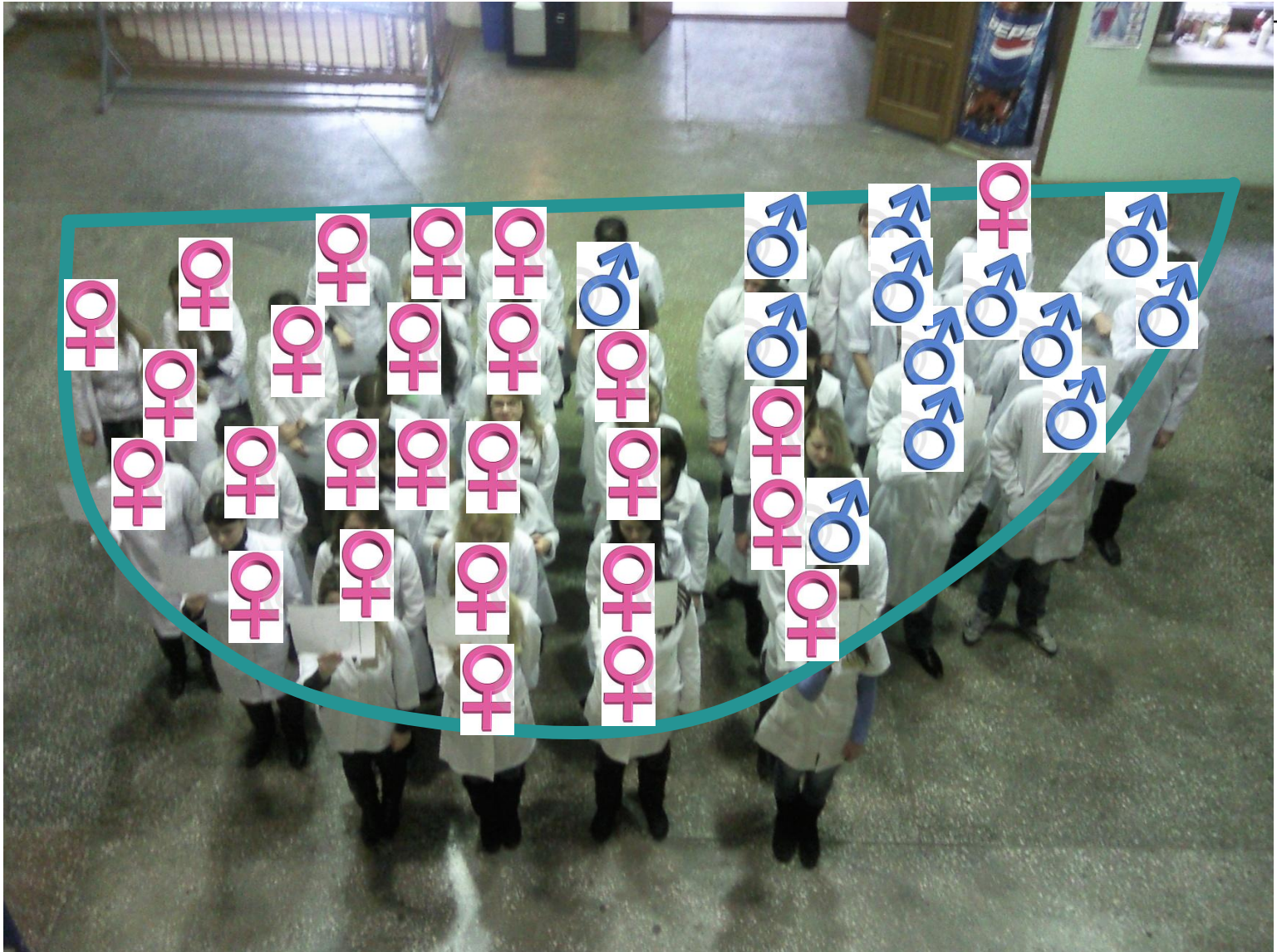
Вид распределения

- соответствие, устанавливаемое между всеми возможными числовыми значениями случайной величины и вероятностями их появления в совокупности

Кривая нормального распределения

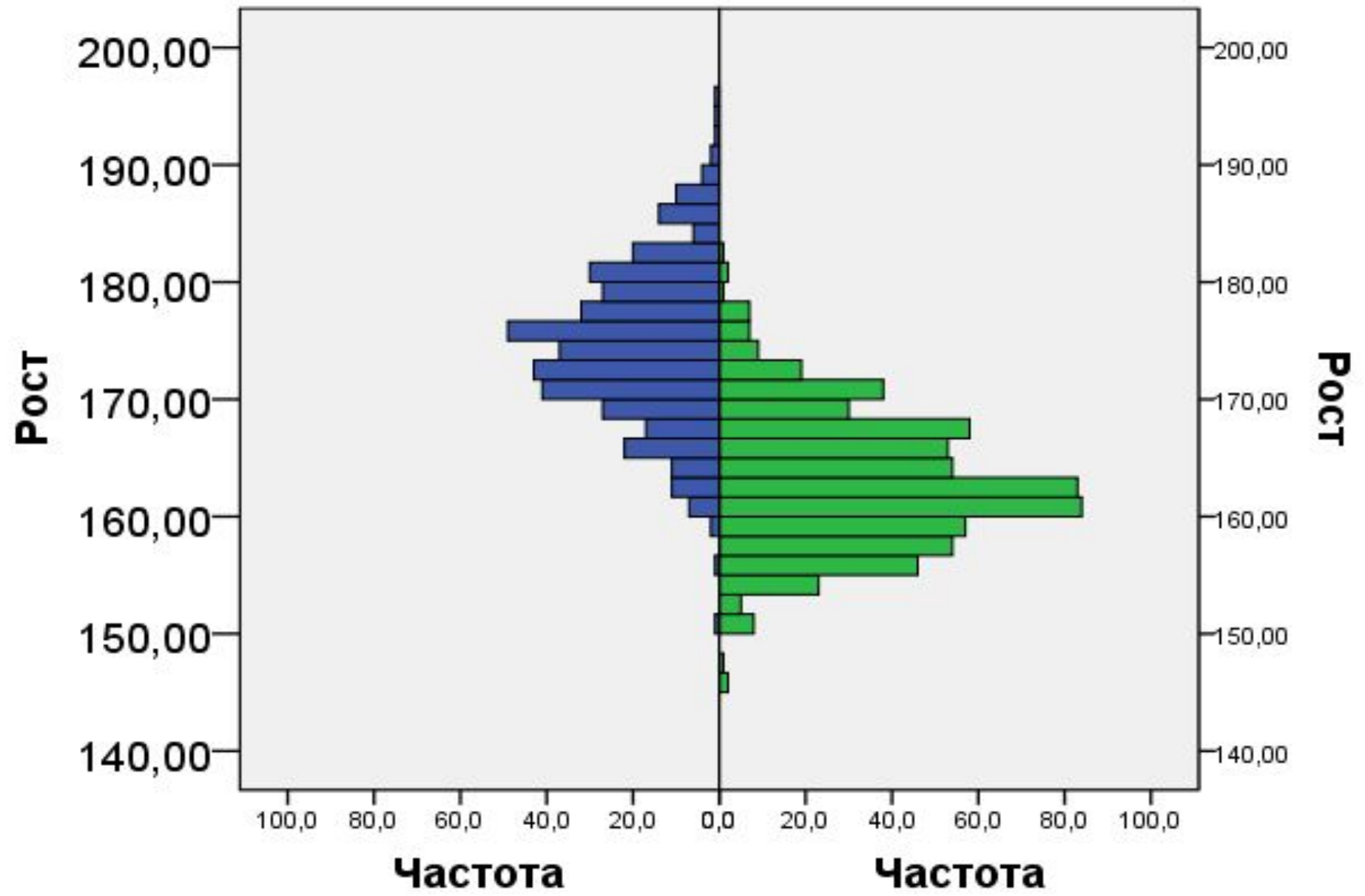
Нормальное (гауссово, симметричное, колоколообразное) распределение – описывает совместное воздействие на изучаемое явление случайно сочетающихся факторов, число которых неограничено велико. Характеризует распределение непрерывных случайных величин.





Пол

Мужской**Женский**



Средняя величина –

- это обобщающий показатель статистической совокупности, который **погашает индивидуальные различия значений** статистических величин, позволяя **сравнивать** разные совокупности между собой.

Среднее арифметическое

Среднее арифметическое n значений обозначают M и определяют как

$$M = \frac{\sum v\rho}{n}$$

Мода (M_o) (mode)- наиболее часто встречающаяся в вариационном ряду варианта.

Мода используется для

дискретных величин:

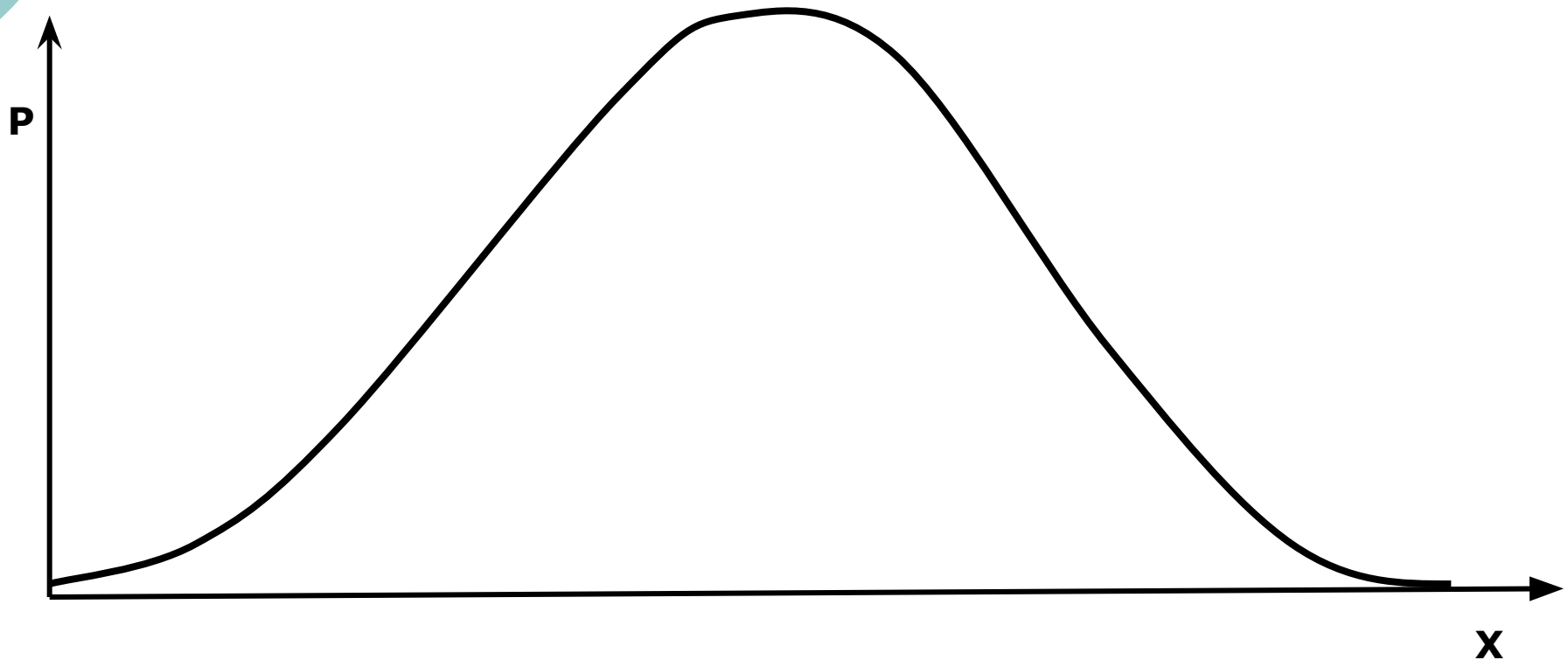
- - при малом числе наблюдений, когда велико влияние состава совокупности на среднюю ;
- - для характеристики центральной тенденции при ассиметричных распределениях, когда велико влияние на среднюю крайних вариантов;

Пример использования моды:

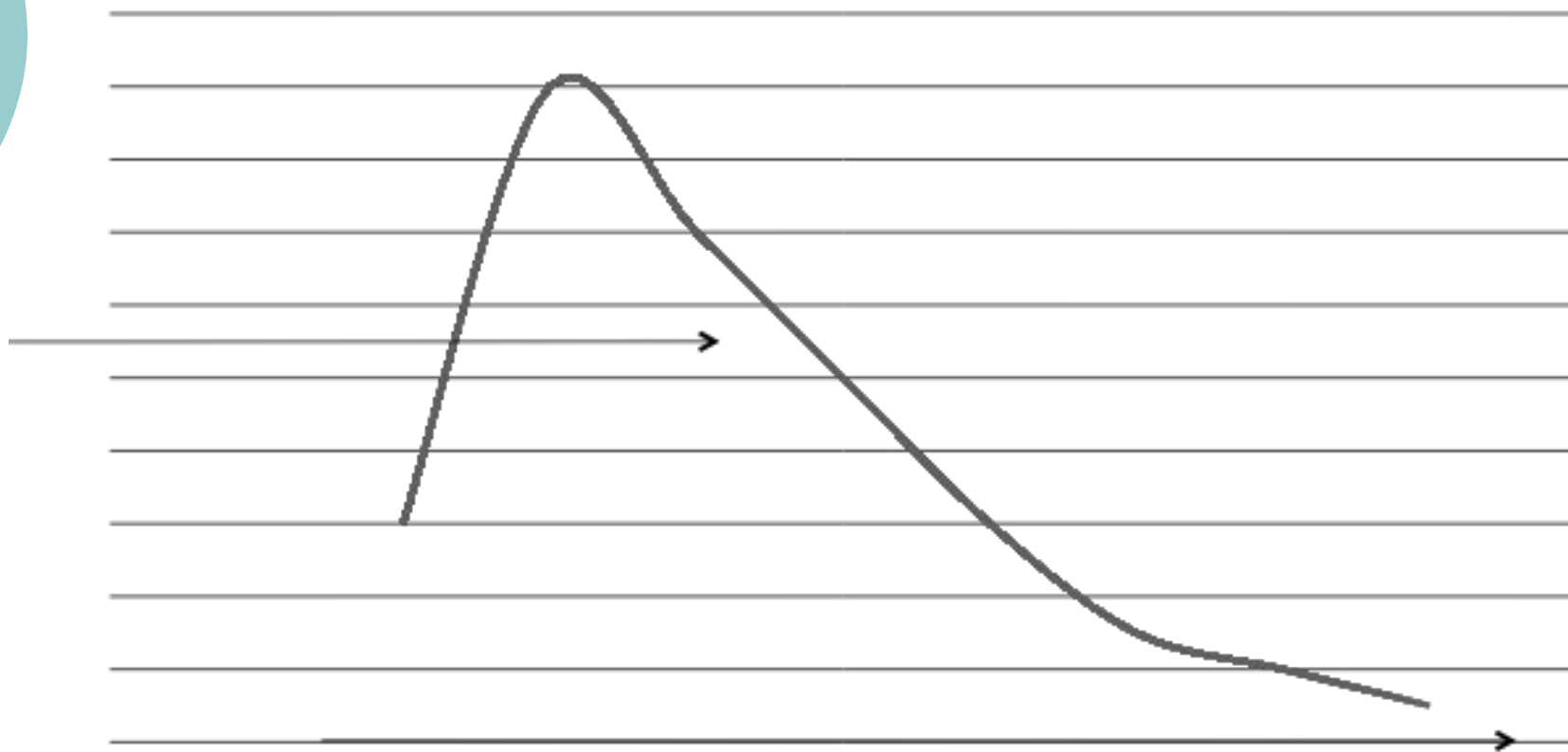
требуется определить среднюю длительность госпитализации рабочих промышленных предприятий в связи с производственным травматизмом

Число госпитализации	дней	2	3	4	5	6	7	8	9	Итого
Число рабочих		6	18	12	10	6	3	2	1	60

Кривая нормального распределения



Ассиметричное распределение



Медиана (Me)(median) -варианта, которая делит вариационный ряд на две равные части.

Медиана используется:

- - при необходимости знать, какая часть вариантов лежит выше и ниже срединного значения ;
- - для характеристики центральной тенденции при ассиметричных распределениях .
- Форма отображения – $Me (V_{25}; V_{75})$

Вариационный ряд можно разбивать на отдельные (по возможности равные) части, которые называются квантилями (quantile). Наиболее часто употребляемые квантили:

**Название квантилей
Число частей, на которые
разбивается ряд**

Медиана

- 2 части

Квартиль

- 4 части

- Квартили определяют, после определения медианы как середины каждой из половин ряда

156, 158, **159**, 162, 166, **167**, 168, 170, **172**, 172, 174

Медиана – 167,

Квартили – 159 и 172

167 (159; 172)

Медиана и квартили

Например, исследуемый признак – «срок, в котором ребенок начал самостоятельно ходить» - в исследуемой группе имеет ассиметричное распределение. При этом, нижнему квартилю (V_{25}) соответствует срок начала ходьбы – 9,5 месяцев, медиане – 11 месяцев, верхнему квартилю (V_{75}) – 12 месяцев.

- Соответственно, характеристика средней тенденции указанного признака будет представлена, как **11 (9,5; 12)** месяцев.

Среднее квадратическое отклонение

- Наиболее полную характеристику разнообразия признака в статистической совокупности дает **среднее квадратическое отклонение** (сигма), которое является общей мерой отклонения вариантов от своей средней величины.
- Форма отображения $M \pm \sigma$

Среднее квадратическое отклонение

$$\sigma = \sqrt{\sum_{i=1}^n \frac{(v_i - M)^2}{n}} = \sqrt{\frac{(v_1 - M)^2 + (v_2 - M)^2 + \dots + (v_n - M)^2}{n}}$$

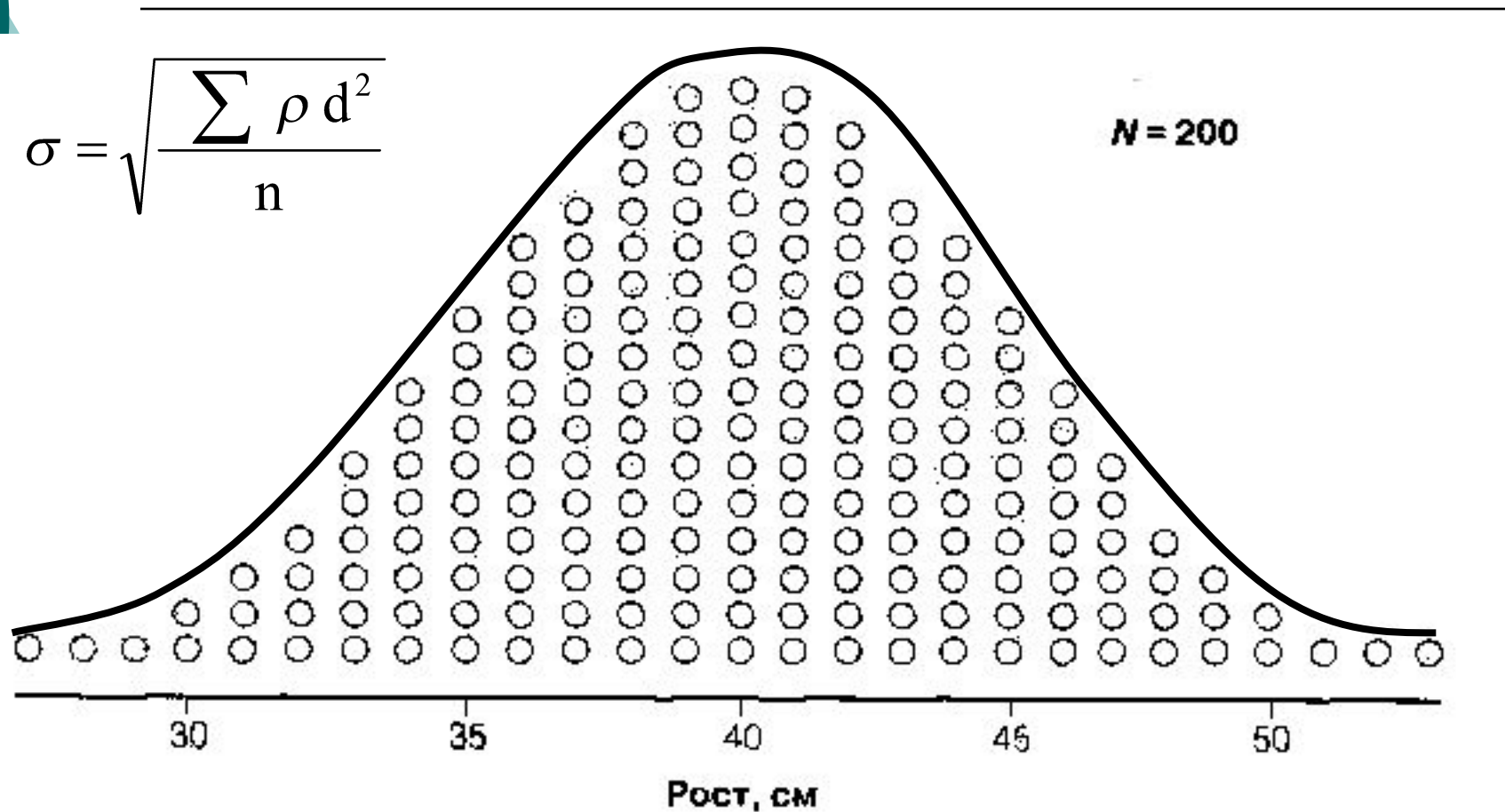
V – варианта

M – средняя арифметическая

P – частота встречаемости варианты

n – число вариантов

Основные характеристики (параметры) нормального распределения



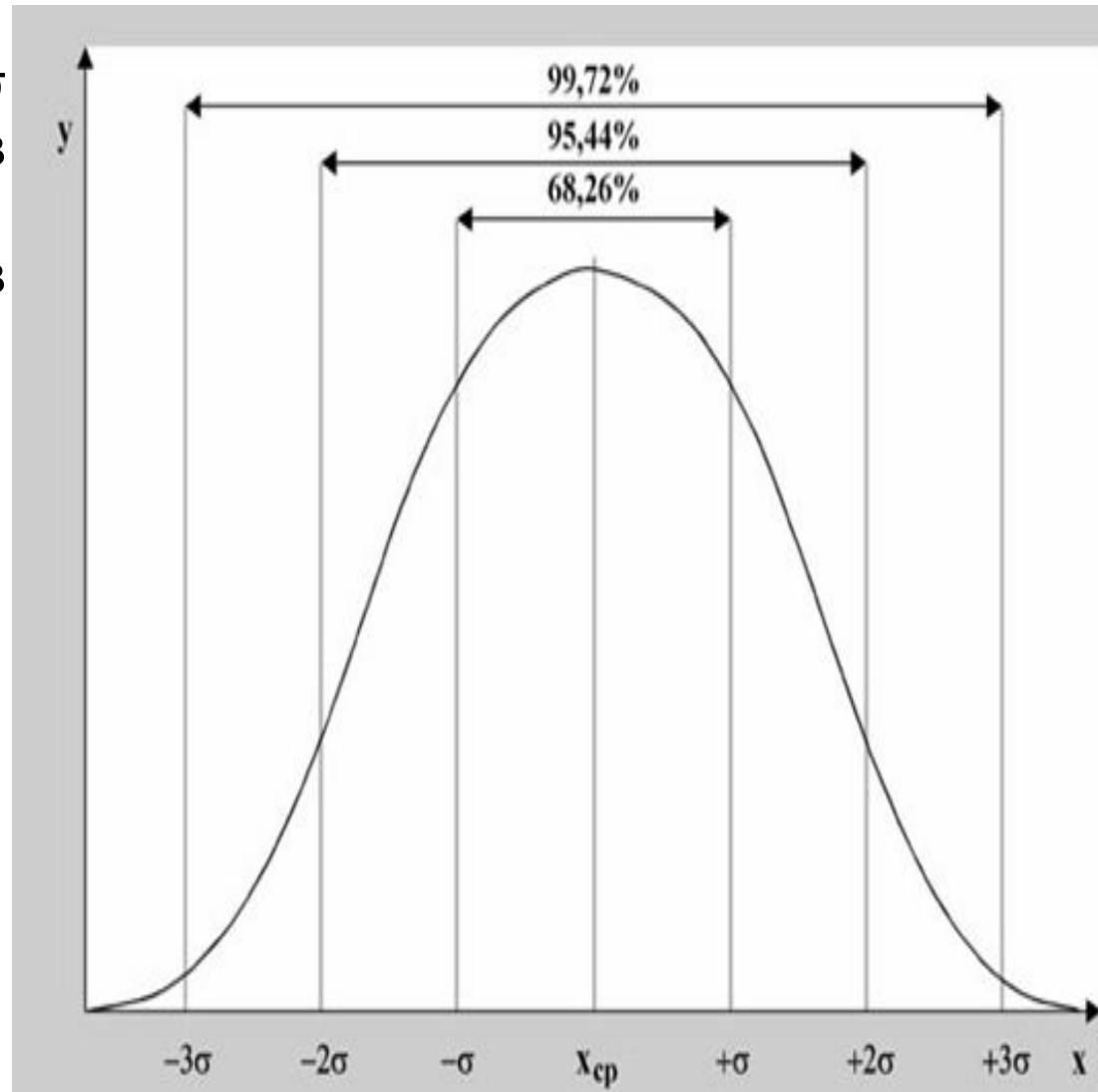
Среднее арифметическое значение (M)

Стандартное (среднеквадратическое) отклонение (σ)

Количество наблюдение (n)

68,3 % всех вариант отклоняются от своей средней не более, чем на σ
95,4% вариант находятся в пределах $X \pm 2\sigma$
99,7% вариант находятся в пределах $X \pm 3\sigma$.

Отклонение параметра от его средней арифметической в пределах σ расценивается как норма, субнормальным считается отклонение в пределах $\pm 2\sigma$ и патологическим - сверх этого предела, т.е. $> \pm 2\sigma$ »



Правило «трех сигм»

Коэффициент вариации

- отношение среднего квадратического отклонения к средней величине признака. Выражается в процентах.

$$V_{Am} = \frac{\sigma}{M} \cdot 100\%$$

Коэффициент вариации

- градации степени разнообразия признака:
 - - слабое — до 10 %
 - - среднее — 10 - 20 %
 - - сильное — более 20 %
- Совокупность считается однородной, если коэффициент вариации не превышает **33%**

Состав работников промышленного предприятия

Учетный признак	Среднее арифметическое	Среднее квадратическое отклонение σ	Коэффициент вариации, %
Стаж работы (лет)	8,7	2,8	32,1
Возраст (лет)	37,2	4,1	11,0
Образование (классов)	9,2	1,1	11,9

Стандартная ошибка

- Для средней:

$$m_p = \frac{\sigma}{\sqrt{n}}$$

- Для доли:

$$m = \sqrt{\frac{P(100 - P)}{n}}$$

Например, процент выздоровления среди больных составил (95,2 ± 2,5)%.

В малых выборках – $n-1$

Доверительный интервал


- Средние значения отражают только уровень признака в выборке!!!
- Как перенести эти данные на генеральную совокупность?

Доверительный интервал

- вокруг выборочного среднего значения строится интервал, который бы с заданной вероятностью – **доверительной вероятностью** – «накрывал» бы истинное значение этого параметра в генеральной совокупности. Этот интервал называется **доверительным интервалом**.

Доверительный интервал

- ⊗ Для нормального распределения :
 - 68,3% всех выборочных средних попадают в интервал $M \pm m$
 - 95,5% - в интервал $M \pm 2 \cdot m$
 - 99,7% - в интервал $M \pm 3 \cdot m$
 - t – доверительный критерий
 - $M_{ген} = M_{выб} \pm t m_M$
 - $P_{ген} = P_{выб} \pm t m_P$

- 
-
- **Среднее арифметическое** наиболее точно характеризует центральную тенденцию при нормальном распределении.
 - **Медиана** – при ассиметричном распределении



-
- Для выбора метода, сравнительной статистики необходимо знать распределение переменной.
 - Для этого применяются критерии Колмагорова-Смирнова и Шапиро-Уилкса.

Оценка достоверности (статистической значимости) различий средних величин

Используются параметрические и непараметрические методы.

- В параметрических методах используют параметры распределения (M, σ)
- *В непараметрических методах – производят ранжирование данных в порядке возрастания или убывания.*
- *Параметрические методы зависят от распределения (оно должно быть нормальным), непараметрические – не зависят.*

Сравнительная статистика

Для параметрической оценки достоверности различий 2-х средних величин (M), используют ошибку средней (m). Расчет средней ошибки относительной величины производится по формуле:

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

Критерии Стьюдента (пример параметрического критерия)

Цель: оценка различий между двумя
средними

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

Если критерий t равен 2, различие достоверно и это можно утверждать с вероятностью безошибочного прогноза, равной 95 % (при $t = 3$ и более - с вероятностью безошибочного прогноза - 99 %).

Величина критерия менее 2 свидетельствует о недостоверном различии сравниваемых показателей.

-
- В первом случае вероятность ошибки $p < 0,05$ или $p < 0,01$
 - Во втором $p > 0,05$, т.е. различия не достоверны

Список литературы:

- 1. Гланц С. Медико-биологическая статистика.-М.: Практика, 1999
- 2. Рунион Р. Справочник по непараметрической статистике.- М.: Финансы и статистика,
○ 1982
- 3. Флетчер Р., Флетчер С., Вагнер Э. Клиническая эпидемиология. Основы доказательной
○ медицины.- М.: Медиа Сфера, 1998
- 4. Реброва О. Статистический анализ медицинских данных.-М.: Медиа Сфера, 2002.
- 5. Сергиенко В.И., Бондарева И.Б. Математическая статистика в клинических
○ исследованиях. – Гэотар Медицина, Москва, 2000, 256 с.
- 6. Платонов А.Е. Статистический анализ в медицине и биологии. – Издательство
○ РАМН,
○ Москва, 2000, 51 с.
- 8. Making Sense Of Data. J.H. Abramson. Second edition. OUP, 1994.
- 9. An Introduction to Medical Biostatistics. Martin Bland. Third edition. Oxford Medical
○ Publications, 2000, 405 p.
- 10. Statistics. David Freedman. W.W. Norton & Company. Third edition, 1998, 850 p.