

Типы распределений и их закономерности

Лекция посвящена трём базовым типам распределений, которые используются в биометрии и генетике для анализа количественных и качественных признаков. Каждое распределение имеет свою область применения и математический аппарат.

1. Нормальное распределение.
2. Биномиальное распределение.
3. Распределение Пуассона.

**Уравнение плотности для
нормального распределения
(уравнение Гаусса-Лапласа) имеет
вид:**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Обозначения в уравнении Гаусса-Лапласа

где, $f(x_i)$ – высота подъёма кривой (плотность вероятности для значения x_i);

μ – математическое ожидание;

$\pi = 3,14159$;

e – основание натурального логарифма ($e = 2,71828$);

σ – среднее квадратическое отклонение;

σ^2 – дисперсия;

x_i – варианта.

Иоган Карл Фридрих Гаусс
(1777-1855)



Пьер-Симон Лаплас
(1749-1827)



1. Нормальное распределение

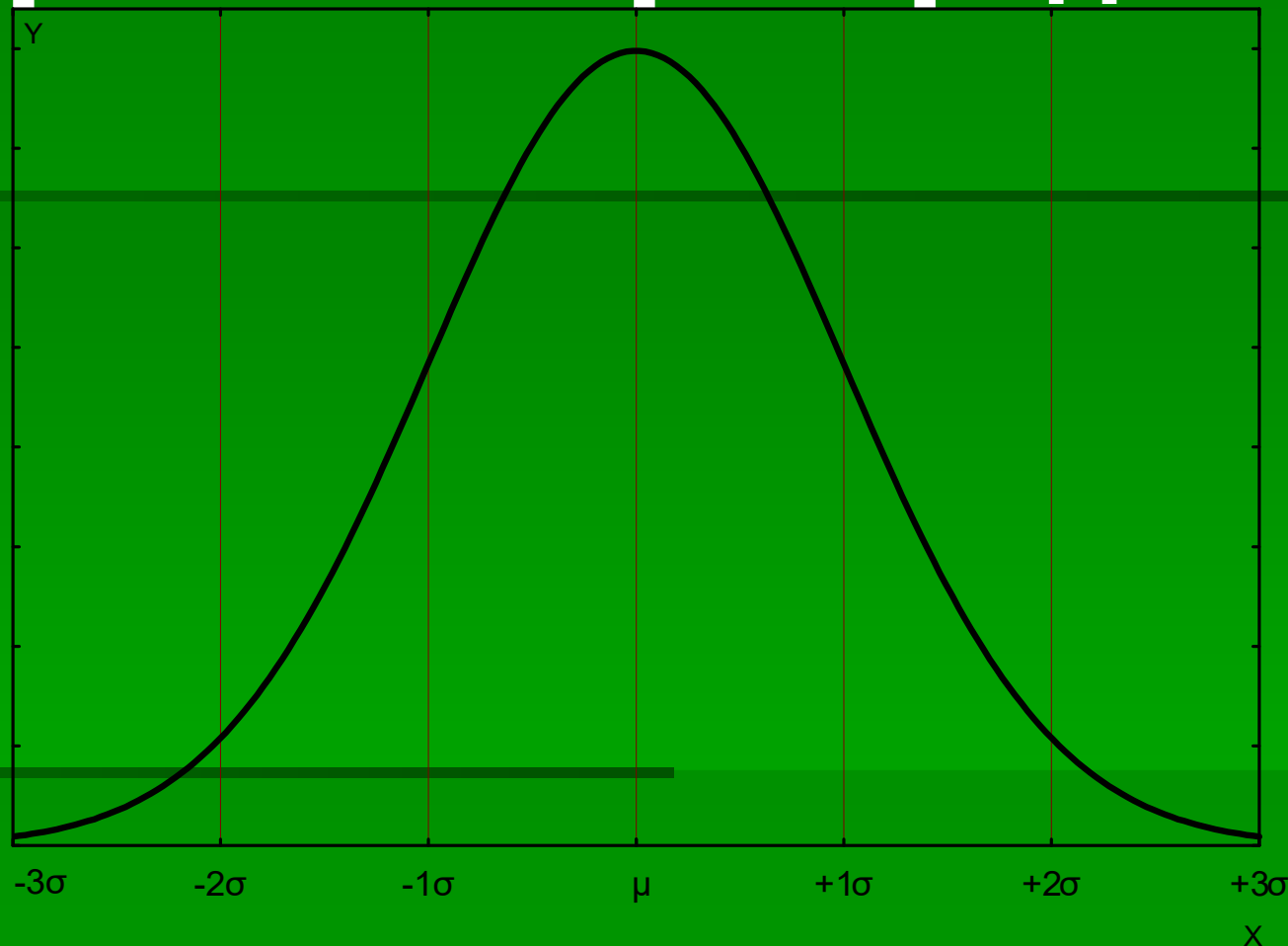


Рис. Кривая нормального распределения.
Отклонения вправо и влево от $\bar{X}$ укладываются в $\pm 3,3 \sigma \sim 99,9\%$.

1. Закономерности нормального распределения

1. Нормальное распределение укладывается в пределах $\pm 3,3 \sigma$ от средней арифметической.

***2. Теоретически показано, что на кривой нормального распределения в пределах $\pm 1\sigma$ укладывается $\sim 68,3\%$ всех вариантов,
в пределах $\pm 2\sigma \sim 95,5\%$,
в пределах $\pm 3\sigma \sim 99,7\%$,
в пределах $\pm 3,3 \sigma \sim 99,9\%$.***

1. Нормальное распределение. Нормированное отклонение

Обозначается показатель латинской буквой t ; часто

называется t – критерием, $t = \frac{x_i - \bar{X}}{\sigma}$, где

x_i - значение i -ой варианты распределения;

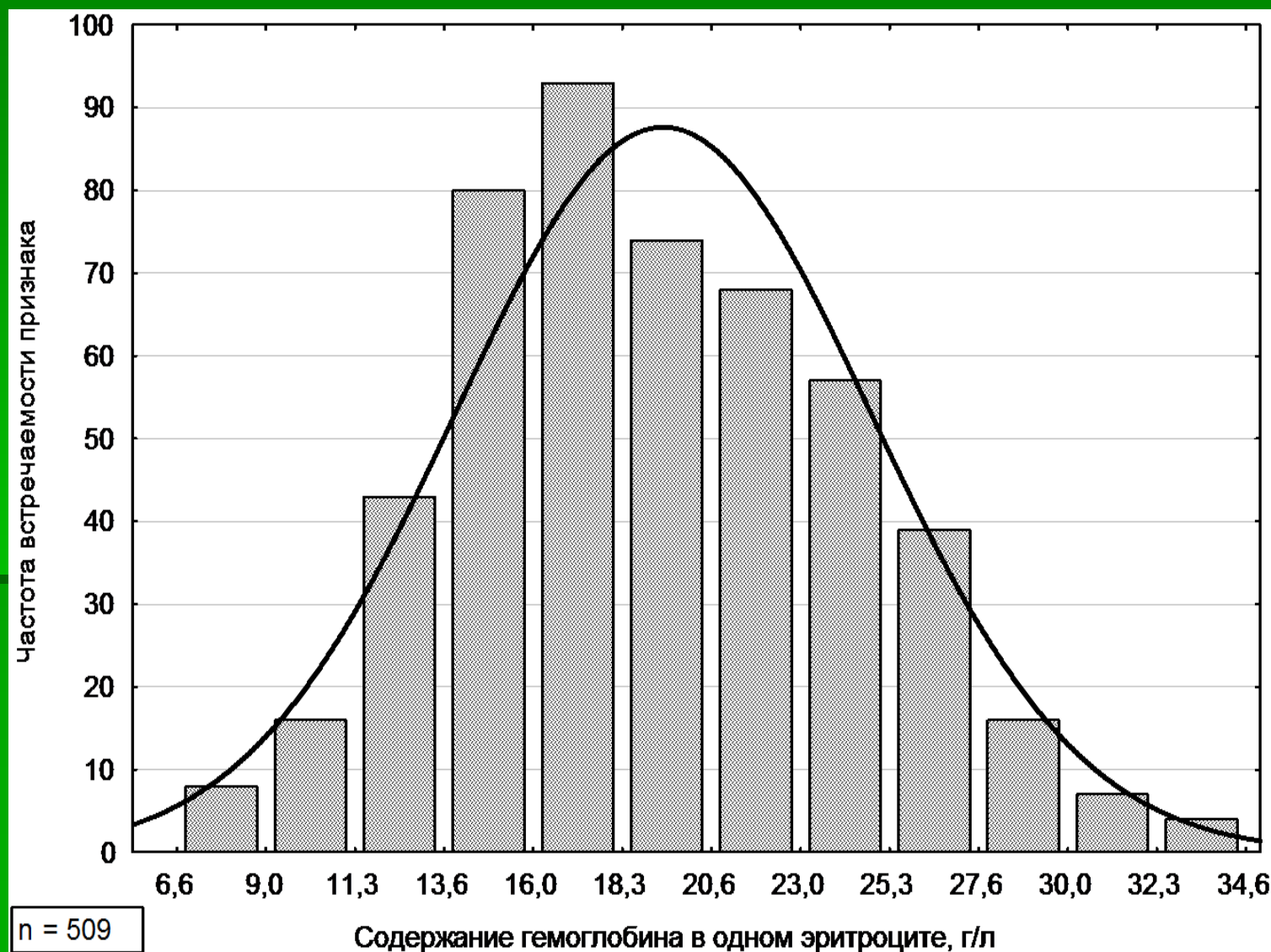
$\bar{X}$ - среднее значение признака в распределении,

σ - среднее квадратическое отклонение, характерное для признака.

Смысл t – критерия состоит в следующем. Если $t = \frac{x_i - \bar{X}}{\sigma}$,

то $x_i - \bar{X} = t \cdot \sigma$.

Гистограмма распределения частот по содержанию гемоглобина в одном эритроците цельной крови поросят-сосунов крупной белой породы



Нормальное распределение

Нормальность распределения можно проверить:

- Зная закономерности нормального распределения (найти $\min$ и $\max$, сигму и определить укладываются фактические границы совокупности в теоретические).
- Анализ графического изображения распределения.
- С помощью критериев, связанных с оценкой нормальности распределений (критерий хи-квадрат, критерий асимметрии и эксцесса, критерий Колмогорова-Смирнова и др.).

2. Биномиальное распределение.

Вариационные ряды, построенные для качественных признаков отличаются от рядов, полученных для количественных. Как правило, альтернативные признаки характеризуются биномиальным типом распределения, которое также называется распределением Бернулли.

2. Биномиальное распределение. **Бином Ньютона**

$$(a + b)^n$$

2. Биномиальное распределение

Вероятность появления члена совокупности с наличием какого-либо признака обозначается буквой p , а его отсутствие – буквой q .

$$(p + q)^n$$

Вероятность

- Вероятность – это возможность получения ожидаемых результатов или возможности приобретения объектом той или иной характеристики в процессе своего формирования.

Вероятность

№ свиноматки	1	2	3	4	5	6	Сумма
Количество поросят, x_i	9	10	11	12	13	14	
Число случаев, P	1	4	18	22	4	1	$n=50$
Вероятность события, $p=P/n$	0,02	0,08	0,36	0,44	0,08	0,02	$\sum p=1$

**При различном количестве событий (n)
распределение может описываться выражением
 $(p+q)^n$.**

Выражение	Число наблюдений в группе (k)
$(p+q)^1=p+q$	$k=1$
$(p+q)^2=p^2+2pq+q^2$	$k=2$
$(p+q)^3=p^3+3p^2q+3pq^2+q^3$	$k=3$
$(p+q)^4=p^4+4p^3q+6p^2q^2+4pq^3+q^4$	$k=4$ и т.д.

Для быстрого получения теоретических частот биномиального ряда можно воспользоваться треугольником Паскаля:

k=1							1		1						
k=2						1		2		1					
k=3					1		3		3		1				
k=4				1		4		6		4		1			
k=5			1		5		10		10		5		1		
k=6		1		6		15		20		15		6		1	
k=7	1		7		21		35		35		21		7		1

3. Распределение Пуассона

- Это распределение относится к дискретной изменчивости, к редким событиям. Такими событиями являются мутации, наследственные дефекты, рождение троен у крупного рогатого скота и т.д.
- Поэтому при пуассоновом распределении значение p очень мало, (так как событие совершается редко), а значение q приближается к единице.

3. Распределение Пуассона.

Формула Пуассона

$$p = \frac{\lambda^m e^{-\lambda}}{m!}$$

3. Распределение Пуассона. Обозначения в формуле Пуассона: где,

n - число появлений редко встречающегося события в n независимых повторных испытаниях;

$e - 2,7183... -$ основание натуральных логарифмов;

$\lambda (\bar{x})$ – среднее число появления редкого события

$$\lambda (\bar{x}) = n \cdot p;$$

$!$ – факториал частоты, или произведение натуральных чисел $(1 \cdot 2 \cdot 3 \cdot ... \cdot m)$.

3. Распределение Пуассона

- Пример. Если в популяции вероятность появления наследственного уродства равна $p=0,002$, то можно определить вероятность появления среди, например, $n=200$ телят 3, 2, 1, 0 животных с уродствами.
- Вероятность рождения, например, 3 животных с уродствами равна:
- Находим среднее число появления животных с уродствами $\lambda (\bar{x}) = n \cdot p = 200 \cdot 0,002 = 0,4$ головы
- $p_{m_3} = (0,4)^3 / 3! \cdot 2,7183^{0,4} = (0,064 / 1 \cdot 2 \cdot 3 \cdot 2,7183^{0,4}) = 0,0072$

Вывод:

Выбор статистического метода напрямую зависит от **типа признака** (количественный или качественный) и **частоты события** (обычное или редкое). Нормальное распределение — базовое для количественных признаков; биномиальное — для альтернативных; распределение Пуассона — для редких событий, включая мутации и наследственные аномалии.