Лекция 3

«Закономерности случайной вариации»

План лекции:

- **Теоретические распределения случайной величины: Биноминальное, нормальное, Пуассона.**
- **Стандартизованные величины: стандартная нормальная кривая. Расчет теоретических частот для эмпирического распределения.**
- **Роль теоретических распределений в биологических исследованиях.**

ЗАКОН НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

- Закон Гауса-Лапласа:

Вероятность отклонения любой варианты (x_i) от центра распределения μ, где x_i-μ=о определяется функцией нормированного отклонения (t)

Математически закон нормального распределения можно выразить формулой Гаусса-Лапласа:

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

где:

- ω (x) - плотность вероятности нормального распределения случайной величины X, имеющей среднее $\mu=0$, и дисперсию $\sigma^2=1$,
- e = 2,718... - основание натуральных логарифмов
- π = 3,14

Нормально распределенная величина

- непрерывная **переменная**,

которая может принимать значения

от $-\infty$ до $+\infty$.

Закон Гауса-Лапласа выражает функциональную связь между вероятностью $P(x_i)$(или $\omega(x_i)$) и нормированным отклонением (t). Графически эта функция выражается в виде кривой вероятности – нормальной кривой.

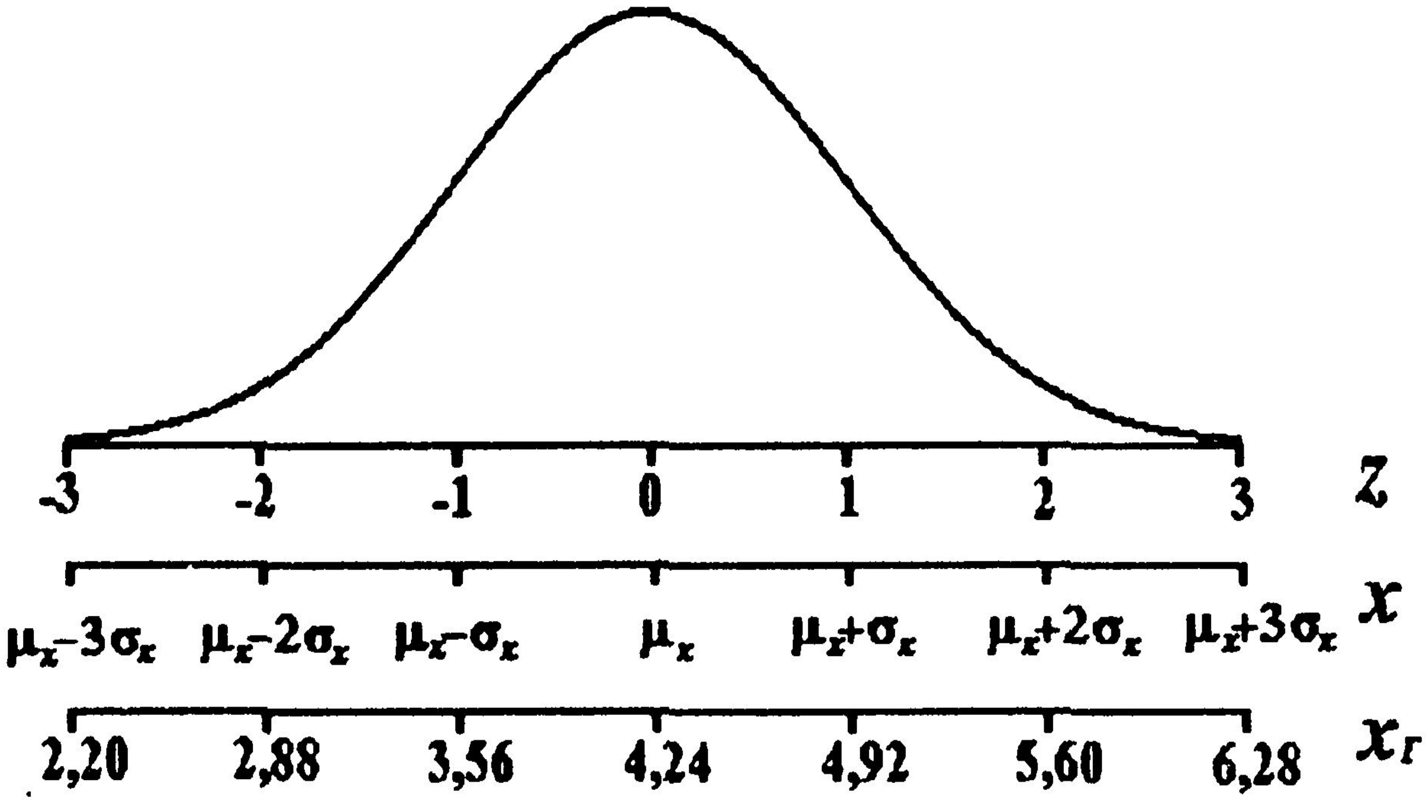


Рис.4.1. Кривая нормального распределения (случайные величины имеют параметры: z: $\mu_z = 0$ и $\sigma_z = 1$; x: μ_x и σ_x; содержание гумуса $x_Г$: $\mu_{x_Г} = 4{,}24$ и $\sigma_{x_Г} = 0{,}68$)

Параметры нормального распределения:

- Средняя величина или математическое ожидание (μ)

$$\mu(x)=x_1p_1 + x_2p_2 + x_3p_3 + \ldots + x_np_n = \sum x_ip_i$$

- Дисперсия случайной величины x – (σ_x^2)

$$(\sigma_x^2)=\mu[x_i-\mu(x)]^2$$

Основные свойства нормального распределения

- 1. Совпадение по абсолютной величине средней арифметической, медианы и моды.
- 2. На равные интервалы, измеряемые нормированным отклонением от центра распределения приходится равное число вариант.

Биноминальное распределение

$$P_n(x) = \frac{n!}{x!(n-x)!} P^x Q^{n-x},$$

Правила сложения и умножения вероятностей

- 1. Вероятность наступления одного из двух (все равно какого) и нескольких независимых и несовместимых событий А, *В*,..., *К* равна сумме их вероятностей:

$$P(A + B + C + \ldots + K) = P(A) + P(B) + P(C) + \ldots + P(K).$$

- 2. Вероятность совместного появления двух или нескольких независимых событий равна произведению вероятностей этих событий:

$$P(A, B, C \ldots K) = P(A) P(B) P(C) \ldots P(K).$$

где знак ! называется факториалом и означает, что следует перемножить целые числа от 1 до числа, стоящего под знаком факториала. Так, $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n$. Заметим, что факториал нуля считается равным единице ($0! = 1$).

Треугольник Паскаля

n	Биноминальные коэффициенты																					2^n
											1											
1										1		1										2
2									1		2		1									4
3								1		3		3		1								8
4							1		4		6		4		1							16
5						1		5		10		10		5		1						32
6					1		6		15		20		15		6		1					64
7				1		7		21		35		35		21		7		1				128
8			1		8		28		56		70		56		28		8		1			256
9		1		9		36		84		126		126		84		36		9		1		512
10	1		10		45		120		210		252		210		120		45		10		1	1024
											и т.д.											

Характер биномиальной кривой определяется двумя величинами:

- числом испытаний;
- вероятностью ожидаемого результата.

РАСПРЕДЕЛЕНИЕ РЕДКИХ СОБЫТИЙ (ЗАКОН ПУАССОНА)

- Когда вероятность ожидаемого события исчисляется сотыми и тысячными долями единицы, распределение частоты такого редкого события в *n* независимых испытаний оказывается крайне асимметричным. Распределение частоты таких редких событий описывается формулой Пуассона:

$$P_n(m)\frac{a^m}{m!}e^{-a} = \frac{a^m}{m!\,e^a}$$

- Где: *m* — частота ожидаемого события в *n* независимых испытаний;
- *a* - наивероятнейшая частота редкого события;
- e = 2,7183...— основание натуральных логарифмов;
- *m!* — факториал частоты, или произведение натуральных чисел 1-2-3... *m.*

- По формуле Пуассона определяется вероятность частоты *m* редких событий в серии повторных испытаний.

При p=0,5 биномиальная кривая строго симметрична и по мере числа испытаний приобретает более плавный ход на всем протяжении. Если же $p \neq q$, биномиальная кривая становится асимметричной особенно при увеличении разницы между p и q.

РАСПРЕДЕЛЕНИЕ МАКСВЕЛЛА

- Наряду с описанными здесь типами распределений случайных величин в биологии встречаются не только симметричные, но и асимметричные распределения, которые, однако, не подчиняются закону Пуассона. Одним из таких распределений является распределение Максвелла:

$$P(x) = \frac{2}{\sqrt{2\pi}} \frac{t^2}{a} e^{-\frac{t^2}{2}} dx.$$

a—параметр распределения, определяемый через среднее значение варьирующего признака по формуле $a = 0{,}6267\ x$;

t=xi/a, где x_i — числовые значения случайной величины X;

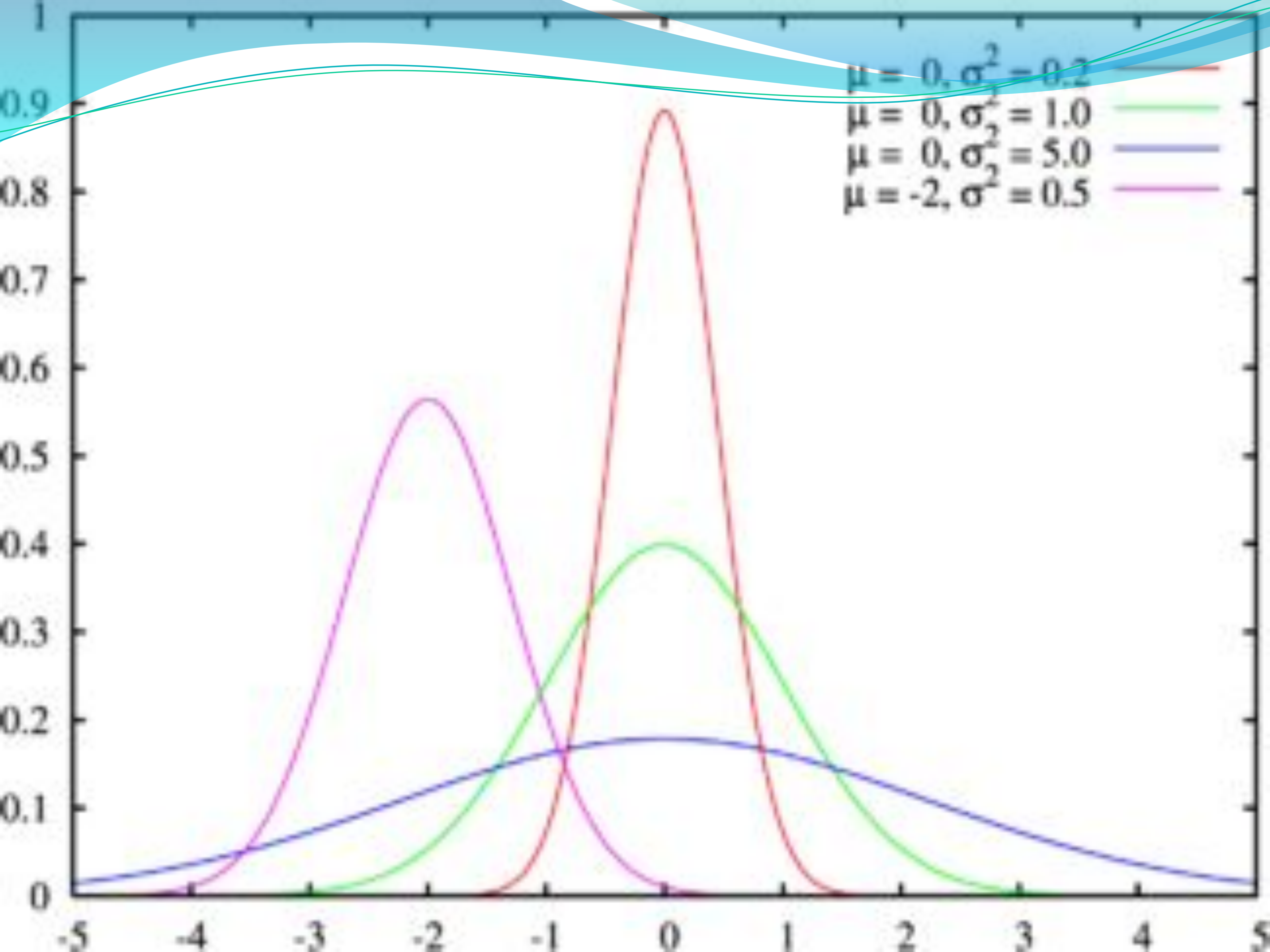
dx-разность между двумя смежными значениями переменной величины X.

Указанием на то, что эмпирическое распределение следует Бэкону Максвелла, служит равенство между средним квадратическим отклонением и величиной 0,674a, т. е. s_x = 0,674a, тогда как распределение Пуассона характеризуется равенством $s_x = x$.

Положение этой кривой полностью определяется двумя параметрами: средней величиной или математическим ожиданием (р) и стандартным отклонением (σ). Если стандартное отклонение σ=1, то нормальная кривая будет иметь стандартную форму, описываемую уравнением:

$$Y=f(t)=\frac{1}{2\pi}e^{-t^2/2}$$

Кривая имеет площадь равную единице. Вершина (y_{max}) соответствует началу прямоугольных координат $xi-\mu=0$.


μ = 0, σ² = 0.2
μ = 0, σ² = 1.0
μ = 0, σ² = 5.0
μ = -2, σ² = 0.5
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-5
-4
-3
-2
-1
0
1
2
3
4
5

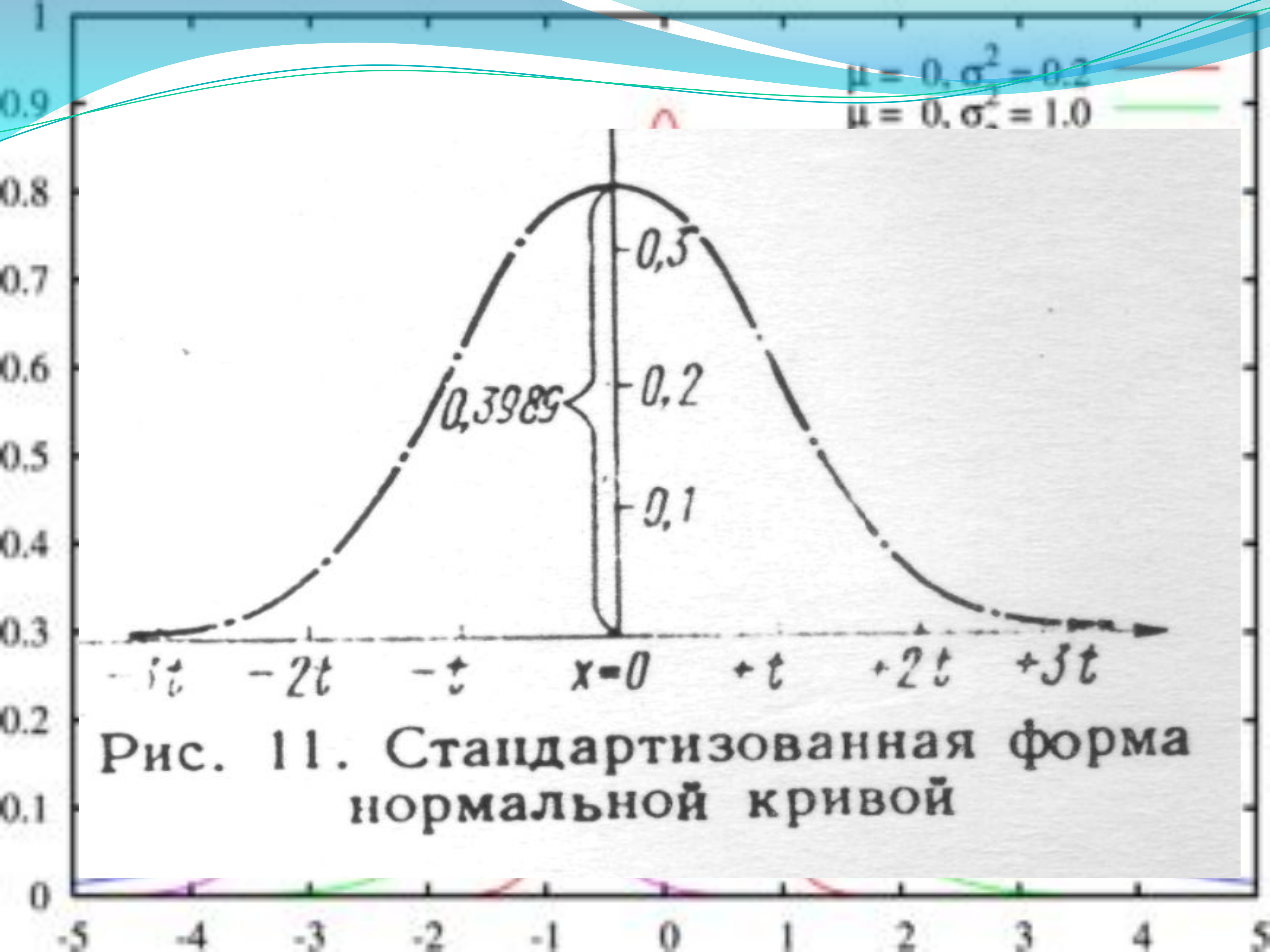


Рис. 11. Стандартизованная форма нормальной кривой

- Положение этой кривой полностью определяется двумя параметрами: средней величиной или математическим ожиданием (р) и стандартным отклонением (σ), характеризующим варьирование отдельных значений случайной величины вокруг центра распределения р. В зависимости от величины о форма нормальной кривой может быть и пологой (при большой величине *а)* и более или менее крутой (при небольшой величине а). Во всех случаях нормальная кривая строго симметрична относительно центра распределения и сохраняет правильную колоколообразную форму. Если стандартное отклонение $\sigma=1$, то нормальная кривая будет иметь стандартную форму. Кривая имеет площадь равную единице. Вершина (y_{max}) соответствует началу прямоугольных координат $x_i-\mu=0$

- Проверка нормальности распределения с помощью показателей асимметрии и эксцесса. Выборочные характеристики — средняя величина и показатели вариации — не содержат информации о законе распределения генеральной совокупности, из которой выборка взята. Трудно судить о законе распределения и по эмпирической вариационной кривой, поскольку на ней сказывается влияние многочисленных случайных причин. Между тем знание закона распределения важно: оно гарантирует от возможных ошибок в оценке генеральных параметров на основании выборочных показателей.
- Многие биологические признаки распределяются нормально. Нередко, однако, эмпирические ряды распределения отклоняются более или менее заметно от нормальной кривой. Эти отклонения могут быть различными, обнаруживая в одних случаях асимметрию, в других — эксцесс, а иногда и то и другое одновременно.
- Асимметрия ряда выражается графически в виде скошенной вариационной кривой, вершина которой может быть сдвинута от центра распределения либо влево, либо вправо. Асимметрию называют *правосторонней* или *положительной,* если вершина кривой сдвинута влево от центра распределения; она более пологая, сильно растянутая по оси абсцисс (рис. 14). При *левосторонней,* или *отрицательной, асимметрии,* наоборот, вершина кривой сдвинута вправо от центра распределения, а ее пологая часть находится на левой стороне (рис. 15).
- Наряду с асимметричными встречаются остро- и плосковершинные кривые распределения. Островершинность вызывается чрезмерным накапливанием численности вариант в центре вариационного ряда, вследствие чего вершина кривой резко поднимается. Кроме одновершинных встречаются двух- и многовершинные распределения.

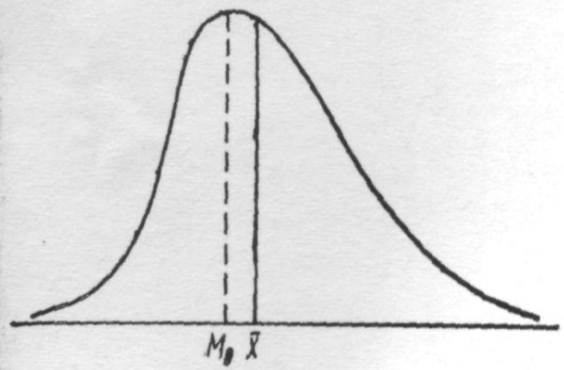


Рис. 14. Асимметричная кривая (положительная асимметрия)

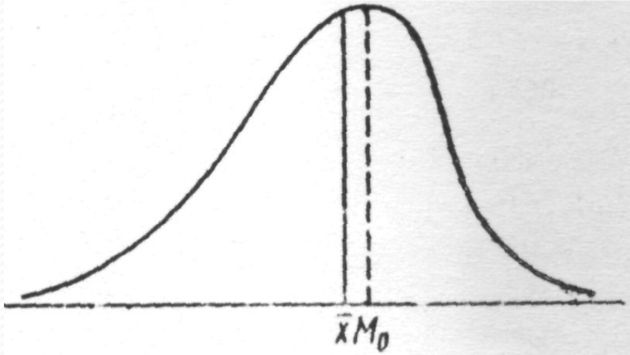


Рис. 15. Асимметричная кривая (отрицательная асимметрия)

Роль теоретических распределений в биологических исследованиях

Законы распределения случайных величин - это вероятностные модели эмпирических распределений. Они служат теоретической основой статистического анализа в самом широком смысле. Различных типов распределений много. Довольно распространенным типом распределения количественных признаков является нормальное распределение. Поэтому нормальный закон особенно важен в биологических (и не только в биологических!) исследованиях; он важен как в теоретическом, так и в прикладном значениях, в частности для выработки нормативов, например, физического развития человека по тем признакам, которые распределяются по нормальном закону или не очень сильно отклоняются от него (метод сигмальных отклонений). Если же эмпирическое распределение не следует нормальному закону и его не удается трансформировать (логарифмированием значений признака) в нормальный ряд, более точными характеристиками при выработке нормативов будут структурные .характеристики - медиана, мода и особенно перцентильные оценки.

Благодарю за внимание!