

федеральное государственное автономное образовательное учреждение высшего образования
Первый Московский государственный медицинский университет им. И.М. Сеченова Министерства здравоохранения Российской
Федерации
(Сеченовский Университет)

Методические рекомендации по дисциплине:

ПРИКЛАДНАЯ БИОСТАТИСТИКА

основная профессиональная образовательная программа
высшего образования - программа специалитета

33.05.01 Фармация

Элементы биостатистики

Ю.С. Жданкина, канд. биол. н., доцент

Кафедра медицинской и биологической физики
Зав. кафедрой д. физ.-мат. н., проф. А.А. Аносов

Сеченовский Университет
Москва, Россия

1. Базовые понятия биостатистики. Основы теории вероятностей.

Ю.С. Жданкина, канд. биол. н., доцент

**Кафедра медицинской и биологической физики
Зав. кафедрой д. физ.-мат. н., проф. А.А. Аносов**

**Сеченовский Университет
Москва, Россия**

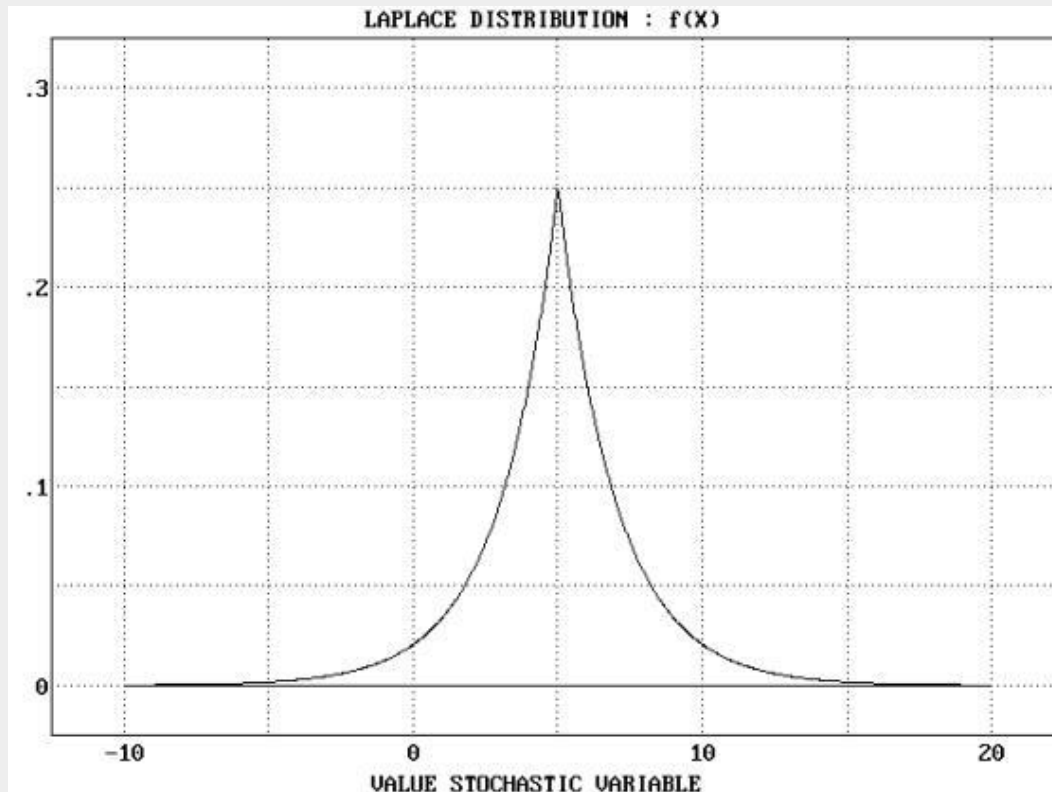
1.1. Введение. История биостатистики.

Ю.С. Жданкина, канд. биол. н., доцент

**Кафедра медицинской и биологической физики
Зав. кафедрой д. физ.-мат. н., проф. А.А. Аносов**

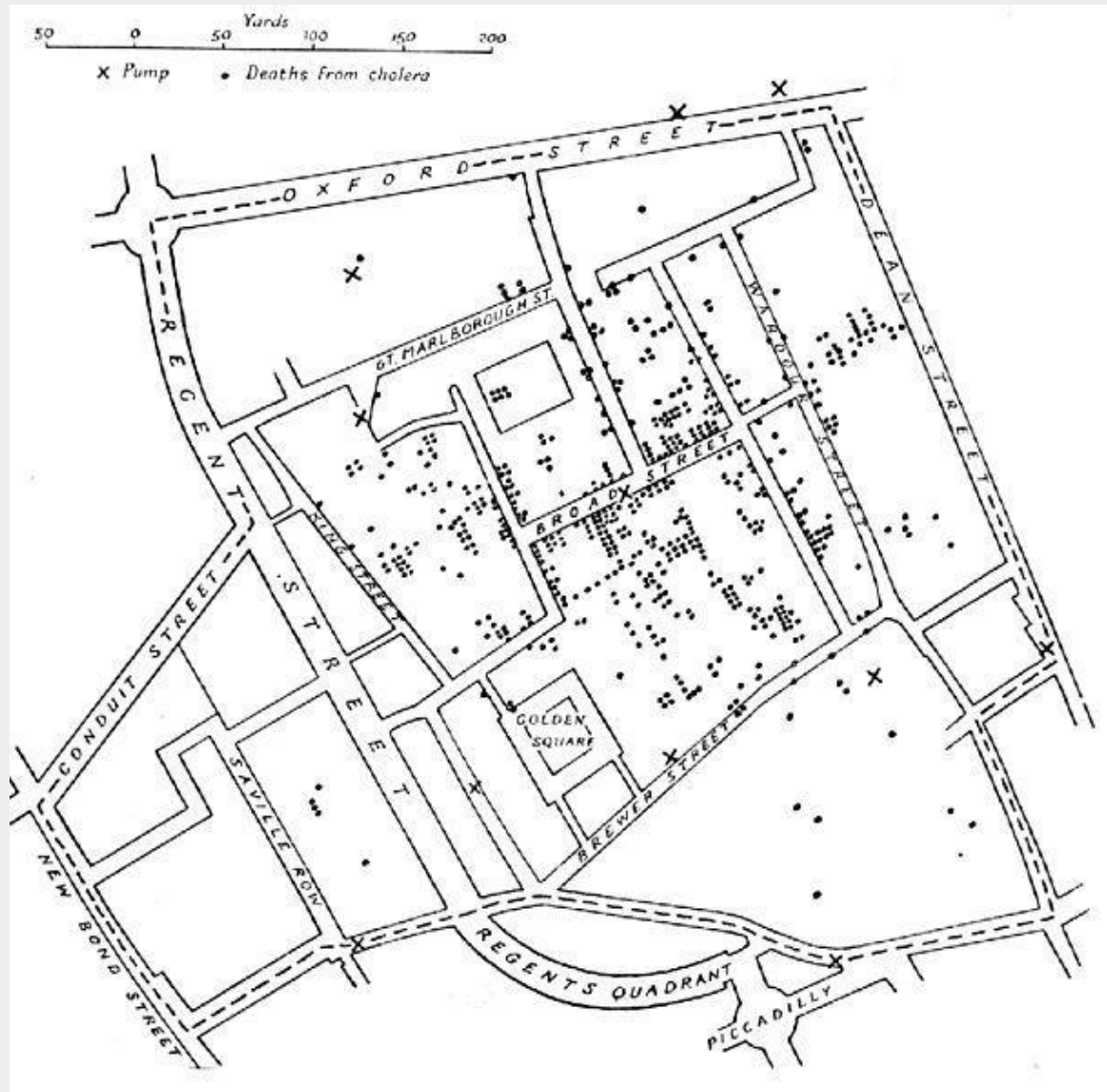
Сеченовский Университет
Москва, Россия

Закон ошибок по Лапласу



$$\varphi(x) = \frac{e^{-|x|}}{2}$$

Джон Сноу. Карта заболевания холерой, 1854 (Возможности статистического графика)



1.2. Группировка результатов наблюдений

Ю.С. Жданкина, канд. биол. н., доцент

**Кафедра медицинской и биологической физики
Зав. кафедрой д. физ.-мат. н., проф. А.А. Аносов**

Сеченовский Университет
Москва, Россия

Пример

Имеется совокупность 20 измерений признака: 2 5 3 6 4 7
4 5 6 6 5 9 5 6 10 8 12 9 7 6

Признак варьирует от 2 до 12 единиц. Расположим эту совокупность в возрастающем порядке: 2 3 4 4 5 5 5 5 6 6 6 6 6 7 7 8 9 9 10 12 – ранжированный ряд значений признака

Учитывая повторяемость величин:

x	2	3	4	5	6	7	8	9	10	11	12
m	1	1	2	4	5	2	1	2	1	1	1

Обозначения

Признаки: $X, Y, Z \dots$

Варианты: $x_1, x_2, x_3 \dots$ или $y_1, y_2, y_3 \dots$

Частоты: m или p

Объем совокупности: n или N

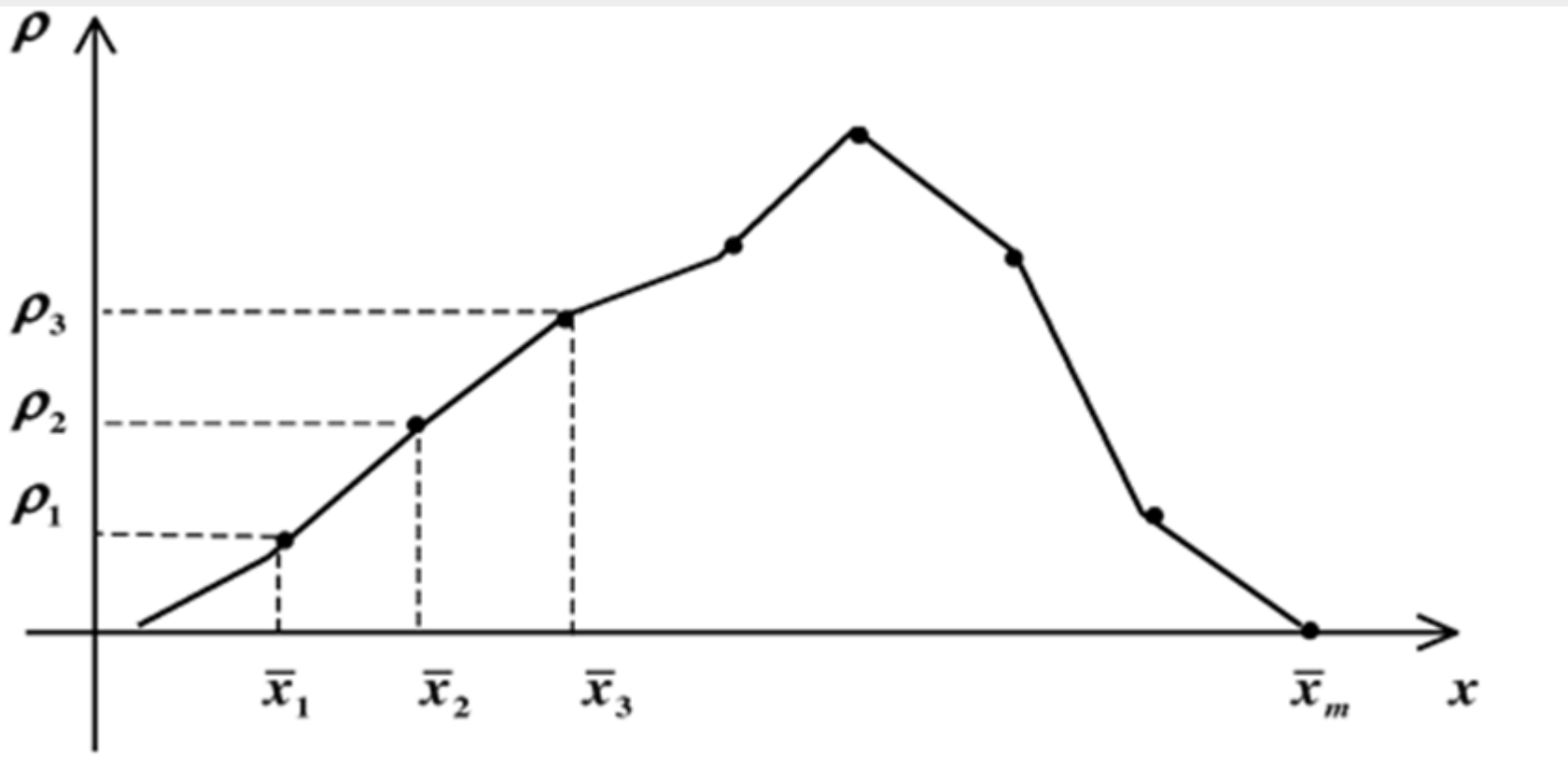
Общая сумма частот равна объему совокупности:

$$\sum p = n \quad (\sum m = n)$$

Относительные частоты:

$$\sum \frac{p}{n} = 1 \text{ или } 100\% \quad (\sum \frac{m}{n} = 1 \text{ или } 100\%)$$

Полигон частот



Интервальные ряды распределения случайной величины

На основании многолетних клинических наблюдений, проводившихся в Сухумском питомнике обезьян, составлена следующая выборка, включающая 100 анализов на содержание кальция (мг%) в сыворотке крови низших обезьян (Павианов-гамадрилов) :

3,6	12,9	12,3	9,9	12,7	11,7	10,8	10,4	10,9	10,2
4,7	10,4	11,6	11,7	12,1	10,9	12,1	9,2	10,7	11,5
3,1	10,9	12,0	11,1	13,5	11,2	13,5	10,1	14,0	10,0
11,6	12,4	11,9	11,4	12,8	11,4	10,9	12,7	13,8	13,2
11,9	10,8	11,0	12,6	10,0	10,3	12,7	11,7	12,1	13,8
12,2	11,9	11,6	10,6	11,1	10,7	12,3	11,5	11,2	11,5
12,7	10,5	11,2	11,9	9,7	13,0	9,6	12,5	11,6	9,0
11,5	12,3	12,8	12,6	12,8	12,5	12,8	11,4	12,5	12,3
14,5	12,3	12,6	11,7	12,2	12,3	11,6	12,0	13,5	12,5
11,6	1,9	12,0	11,4	14,7	11,3	13,2	14,3	13,2	14,2

Интервальные ряды распределения случайной величины

Признак варьирует непрерывно в пределах от 9,0 до 14,7 мг%

$$\alpha = \frac{x_{max} - x_{min}}{k}, \quad k = 1 + 3,32 \lg n, \quad \alpha \approx 0,8$$

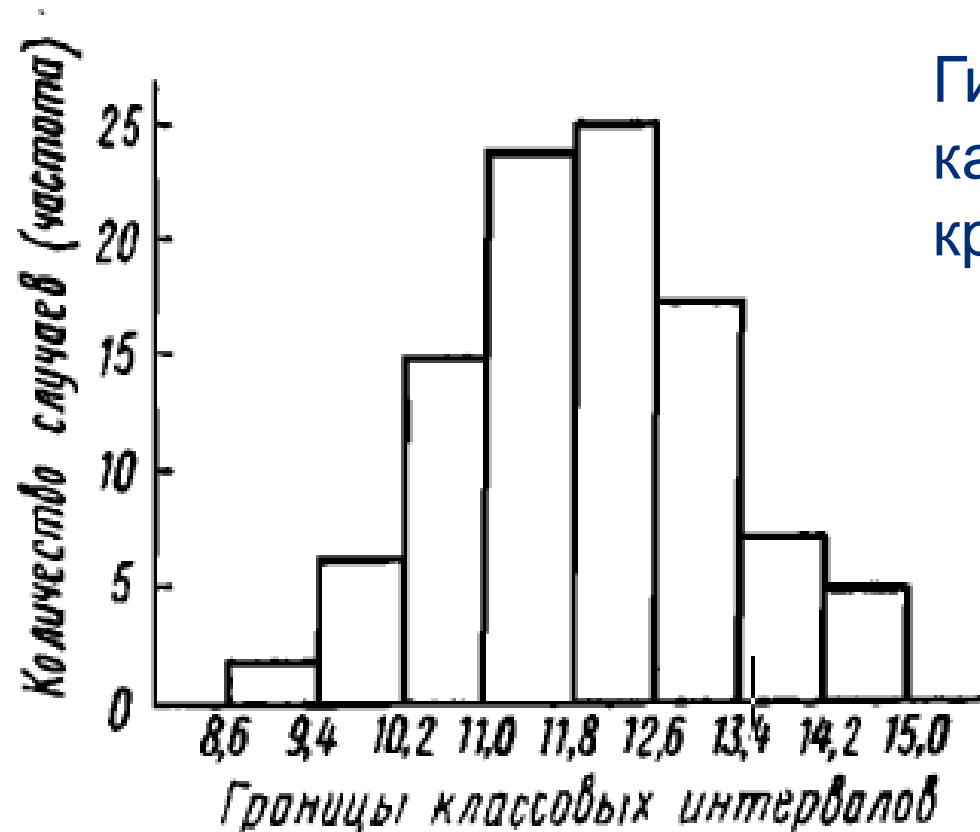
При построении интервального вариационного ряда следует поступать так, чтобы минимальная варианта совокупности попадала примерно в середину первого классового интервала. Выполнение этого требования гарантирует построение вариационного ряда, наиболее полно отвечающего природе изучаемого явления, а следовательно, и наименьшие потери информации о точности вычисляемых статистических характеристик ряда. Этому требованию удовлетворяет формула

$$x_H = x_{min} - \frac{\alpha}{2} = 8,6$$

Затем намечаем следующие классовые интервалы: 8,6-9,4-10,2-11,0-11,8-12,6-13,4-14,2-15,0. Получилось восемь интервалов. Разграничиваем их на величину, равную точности измерения признака, т. е. уменьшаем верхние границы интервалов на 0,1 мг%. Строим вспомогательную расчетную таблицу и разносим все 100 вариант по намеченным классовым интервалам

x	8,6-9,3	9,4-10,1	10,2-10,9	11,00-11,7	11,8-12,5	12,6-13,3	13,4-14,1	14,2-14,9
m	2	6	15	23	25	17	7	5

Гистограмма



Гистограмма распределения
кальция (мг %) в сыворотке
крови обезьян

1.3. Закономерности распределения

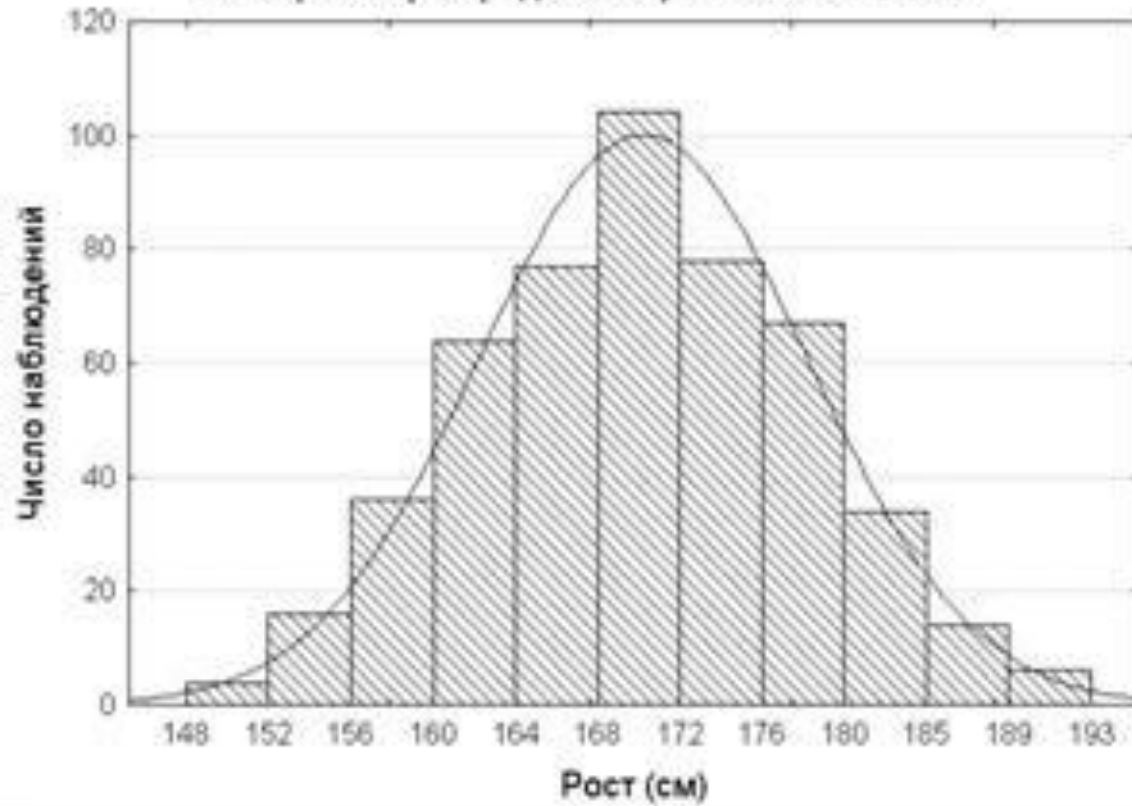
Ю.С. Жданкина, канд. биол. н., доцент

**Кафедра медицинской и биологической физики
Зав. кафедрой д. физ.-мат. н., проф. А.А. Аносов**

**Сеченовский Университет
Москва, Россия**

Закономерности распределения

Гистограмма распределения роста 500 человек



Вероятность и ее свойства

208 мальчиков

200 девочек

408 детей всего

$\frac{208}{408} = 0,51$ - относительная частота родившихся
мальчиков

$\frac{200}{408} = 0,49$ - относительная частота, или доля
родившихся девочек

Вероятность и ее свойства

$$P(A) = \frac{m}{n}$$

Пример: в урне помещается 5 белых и 10 черных шаров. Наугад вынимается один шар. Какова вероятность, что вынутый шар окажется белым?

$$P(A) = \frac{5}{15} = \frac{1}{3} = 0,33$$

Вероятность появления черного шара при однократном тираже: $P(\bar{A}) = \frac{10}{15} = \frac{2}{3} = 0,67$

Свойства вероятности

$$1. 0 \leq P(A) \leq 1$$

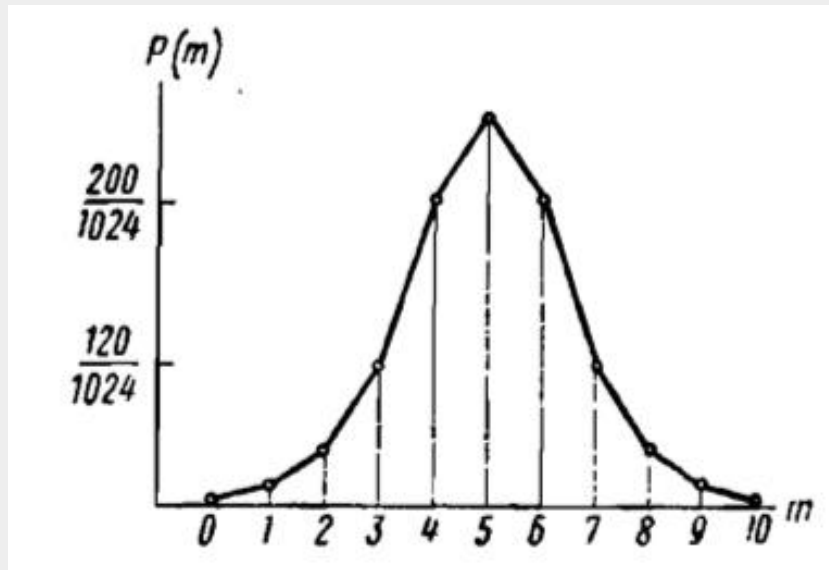
$$2. P(A) + P(\bar{A}) = 1$$

$$P(A) = p, \quad P(\bar{A}) = q$$

$$p + q = 1, \quad p = 1 - q$$

Биномиальное распределение

$$P_n(m) = C_n^m \times p^m \times q^{n-m} = \frac{n!}{m!(n-m)!} \times p^m \times q^{n-m}$$



Кривая распределения вероятностей двучлена $(\frac{1}{2} + \frac{1}{2})^{10}$

Распределение Пуассона

$$P_n(m) \approx \frac{\alpha^m}{m!} \times e^{-\alpha}$$

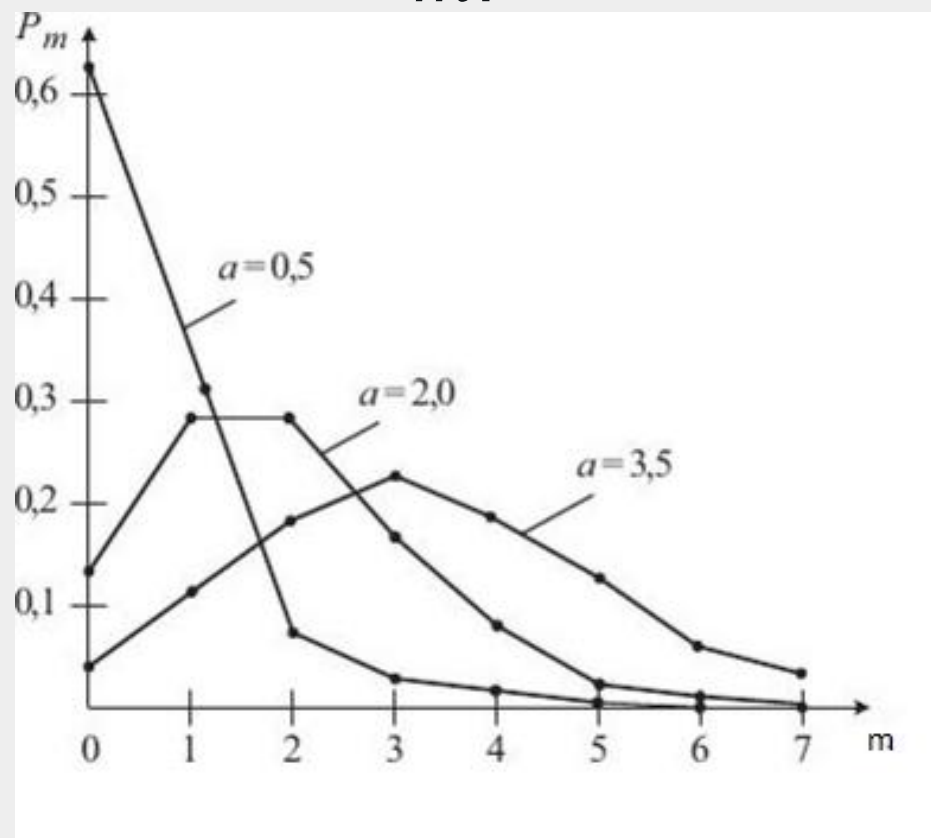


График функции $P_n(m) \approx \frac{\alpha^m}{m!} \times e^{-\alpha}$
для разных значений α

Пример

Вероятность заболевания гриппом после проведения вакцинации для жителей определенного района составляет 0,3%. Определить вероятность того, что из 1000 жителей 10 заболеют гриппом.

Решение:

1. Рассчитаем параметр α :

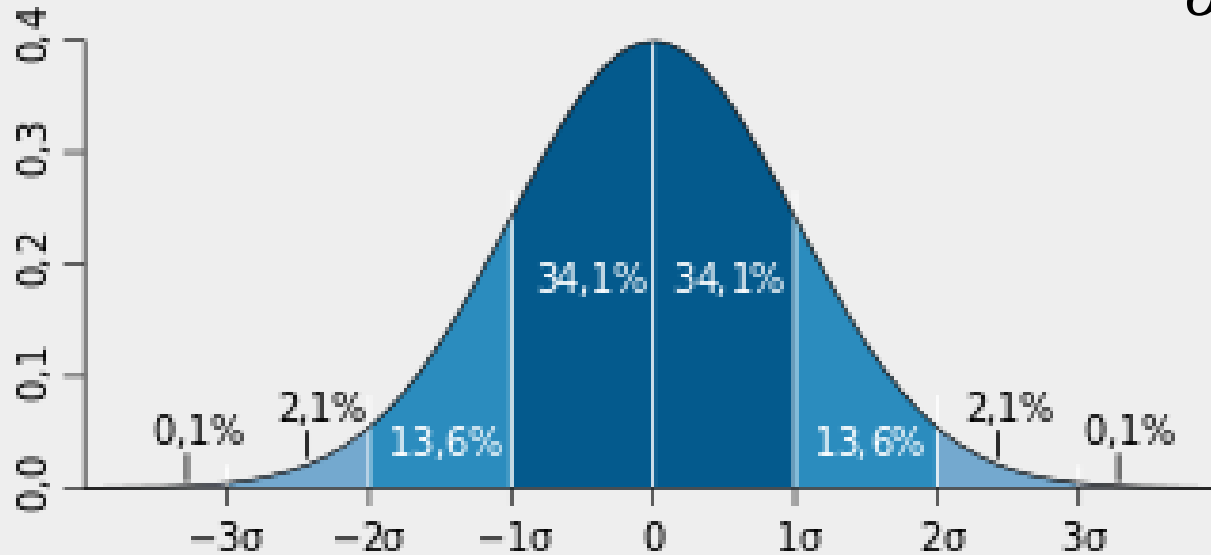
$$\alpha = 0,003 \times 1000 = 3$$

2. По формуле Пуассона:

$$P_{10000}(10) \approx \frac{3^{10}}{10!} \times e^{-3} \approx 0,0008 \approx 0,08\%$$

Нормальное распределение (распределение Гаусса)

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



$$t = \frac{x - \mu}{\sigma}$$

$$f(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-0,5t^2}$$

1.4. Средние величины

Ю.С. Жданкина, канд. биол. н., доцент

Кафедра медицинской и биологической физики
Зав. кафедрой д. физ.-мат. н., проф. А.А. Аносов

Сеченовский Университет
Москва, Россия

Простая средняя арифметическая

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$

Средняя квадратическая

$$\overline{x_q} = \sqrt{\frac{\sum x^2}{n}} \quad \text{или} \quad \overline{x_q} = \sqrt{\frac{\sum px^2}{n}}$$

Непараметрические средние

Медиана (Me) – характеризует середину вариационного ряда

x	3	6	7	8	10	12	13	15	17
p	2	3	4	5	6	5	4	3	2

Me = 10

$$Me = x_k + i\left(\frac{\frac{n}{2} - S_k}{p_k}\right)$$

Непараметрические средние

Мода (Mo) – наиболее часто встречающаяся величина

Классы по уровню кальция в сыворотке крови (мг%)	Частоты (p)
8,55-9,24	2
9,25-9,94	3
9,95-10, 64	9
10,65-11,34	17
11,35-12,04	25
12,05-12,74	23
12,75-13,44	10
13,45-14,14	7
14,15-14,84	4

$$Mo = x_k + i\left(\frac{p_2 - p_1}{2p_2 - p_1 - p_3}\right)$$

Mo=25

1.5. Показатели вариации

Ю.С. Жданкина, канд. биол. н., доцент

Кафедра медицинской и биологической физики
Зав. кафедрой д. физ.-мат. н., проф. А.А. Аносов

Сеченовский Университет
Москва, Россия

Дисперсия и среднее квадратическое отклонение

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

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

Коэффициент вариации

$$CV = 100 \frac{\sigma}{\bar{x}} \%$$

$$CV_1 = 100 \frac{1,2}{11,9} = 10,1\%$$

$$CV_2 = 100 \frac{1,35}{114,74} = 1,2 \%$$

Нормированное отклонение

$$t = \frac{x - \bar{x}}{\sigma}$$

$$t = \frac{171,2 - 164,8}{5,8} = +1,1$$

$$t_1 = \frac{3,31 - 3,10}{0,9} = 0,233, \text{ или } 23,3\%$$

$$t_2 = \frac{2,70 - 2,63}{0,3} = 0,233, \text{ или } 23,3\%$$