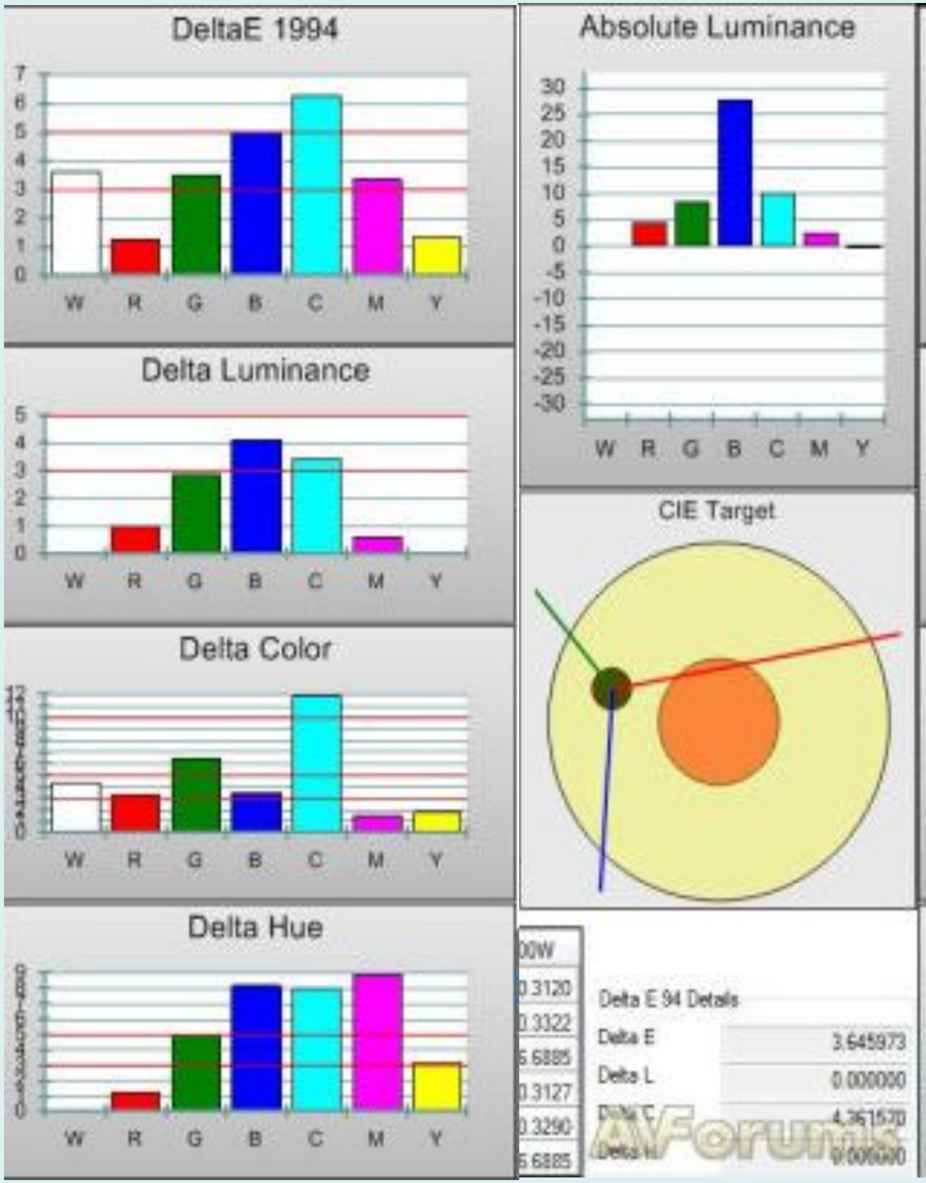
Описание случайных погрешностей с помощью методов математической статистики



43,8% всех статистических данных бесполезны.

Мудрость аналитика

Основные задачи математической статистики

- Описание выборочных данных
- Оценивание (вероятностное) параметров распределения
- Проверка статистических гипотез о свойствах генеральной совокупности

Основные понятия математической статистики

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА — наука о методах обработки экспериментальных данных, полученных при изучении закономерностей в массовых измерениях, выполненных при одинаковых условиях.

СЛУЧАЙНАЯ ВЕЛИЧИНА (X) — физическая величина, значение которой при измерении изменяется от к случаю с той или иной вероятностью.

Дискретная случайная величина - величина, которая принимает отдельные, изолированные значения с определенными вероятностями.

Непрерывная случайная величина - величина, которая может принимать все значения из некоторого конечного или бесконечного промежутка.

ВЫБОРКА ОБЪЕМА (n) — конечная совокупность значений $X = (x_1, x_2, x_3 \ldots x_n)$ полученных в результате n независимых экспериментов, выполненных при одинаковых условиях.

ГЕНЕРАЛЬНОЙ СОВОКУПНОСТЬ — множество всех мыслимых значений случайной величины X.

СТАТИСТИЧЕСКИЕ МОДЕЛИ — зависимости между входными и выходными переменными, носящие вероятностный характер, например:

$$\boxed{x_i = x_{\text{ист}} + \Theta + \delta}$$

Основные понятия математической статистики

ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ X называется функция F(x), определяющая для каждого значения x, вероятность того, что случайная величина X примет значение меньше x. При соблюдении известных условий полностью определяет случайную величину.

ВЕРОЯТНОСТЬ — категория, обозначающая количественную степень возможности появления массовых случайных событий при фиксированных условиях наблюдения, характеризующую устойчивость их относительных частот

$$P(X) = n/N$$

Функция распределения

Интеграль ... ость
того, что сл ... ия,
меньшие н

(1)

Свойства ф

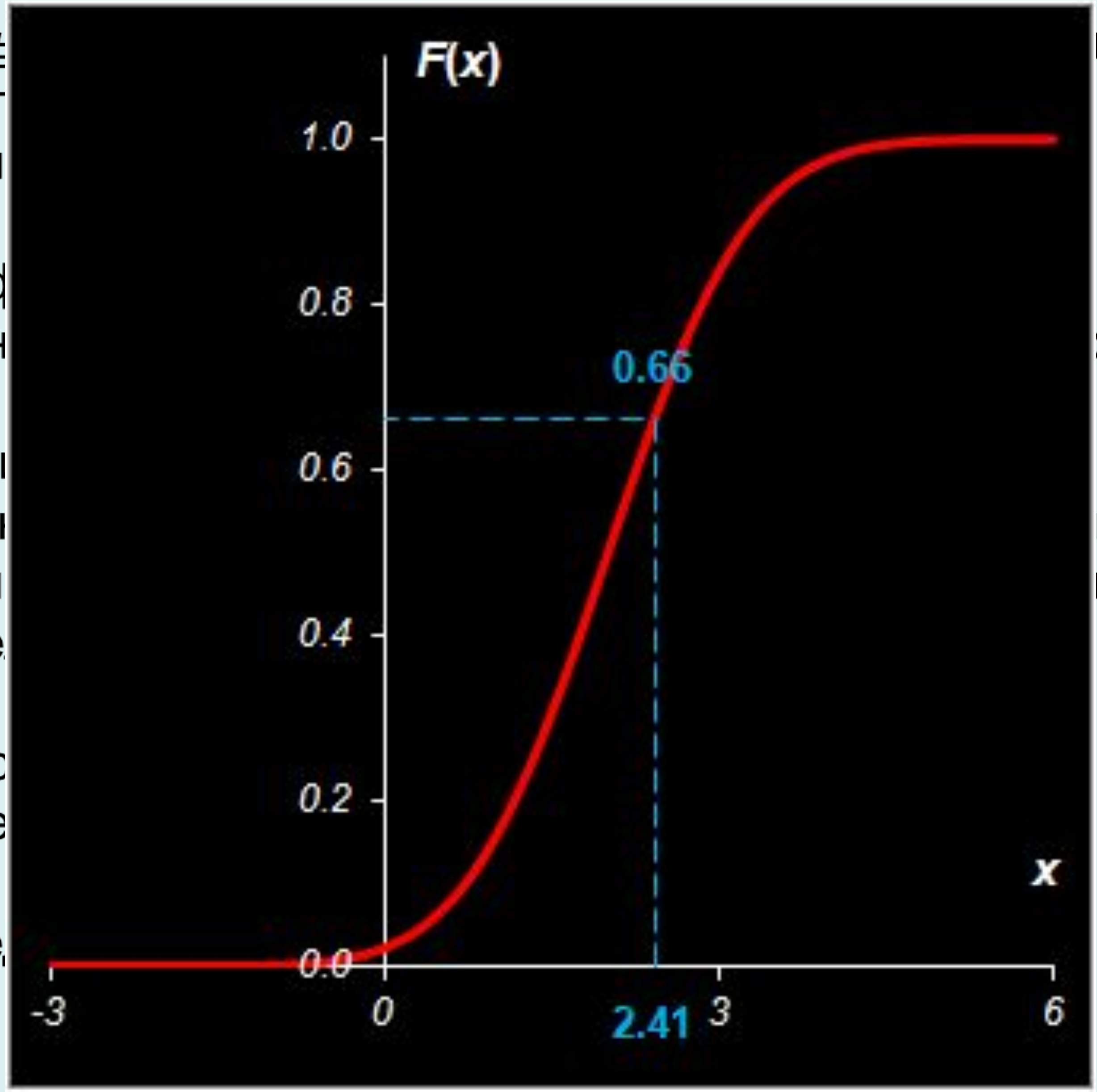
❖ Значен ... зку

❖Функци

❖Вероят ... начение,
заключен ... кции
распреде

❖Если вс
принадле

❖Справе



Функция распределения

Дифференц
плотности функции.

$\varphi(x)$ (2)

Свойства пл

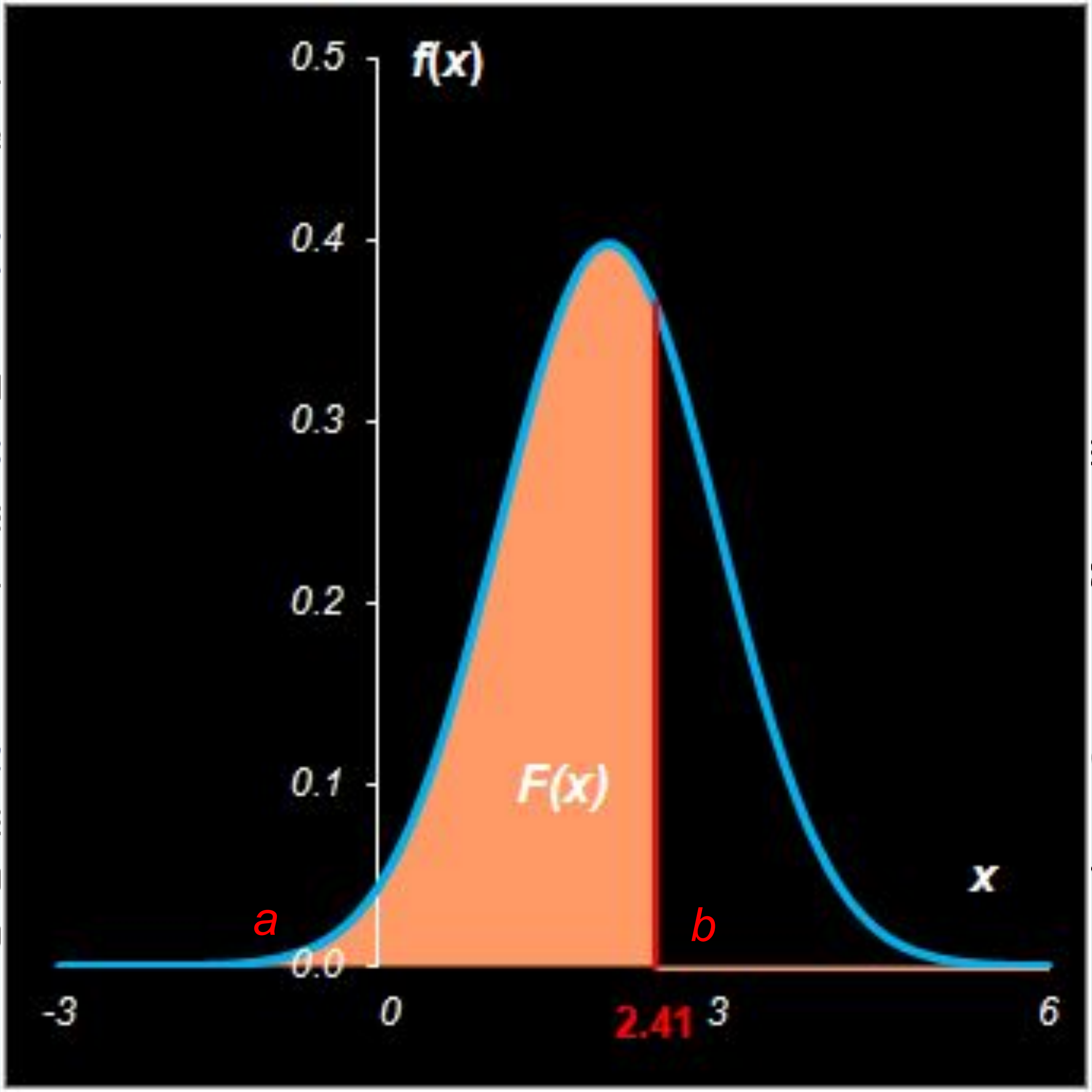
- $\int_{-\infty}^{+\infty} \varphi($ вна 1,

иными слов
дифференц ; равна
единице;

- $\int_{a}^{b} \varphi(x)$ опадания

случайной
От диффер ти к
интегральн

(3)



Важнейшие параметры функции распределения

- ***Момент случайной величины –*** числовая характеристика распределения данной случайной величины.

Начальный момент k-го порядка :

$$\mu_k = \int_{-\infty}^{+\infty} x^k \varphi(x)\,dx; \qquad \mu_k = \sum_{i=1}^{n} x_i^k p_i.$$

где p_i – вероятность появления дискретной величины. Здесь и ниже первая формула относится к непрерывным, а вторая к дискретным случайным величинам.

По определению μ_1 = M(X) и показывает относительное расположение распределения на числовой прямой.

- ***Математическое ожидание M(X)*** - генеральное среднее.

Для непрерывной неограниченной, находящейся в интервале от –∞, до +∞, или ограниченной интервалом [a, b] случайной величины

- определяется интегралом

$$M(X) = \int_{-\infty(a)}^{+\infty(b)} x\varphi(x)\,dx \qquad (4)$$

- Для дискретной случайной величины: $M(X) = \sum_{i=1}^{n} x_i P_i$ (5)

Важнейшие параметры функции распределения

- Центральные моменты k-го порядка :

$$\mu_k = \int_{-\infty}^{+\infty} (x - m_1)^k \varphi(x) dx; \qquad \mu_k = \sum_{i=1}^{n} (x_i - m_1)^k p_i.$$

По определению μ_2= D(X) и показывает разброс распределения вокруг среднего значения.

- ***Дисперсией D(X)*** случайной величины *X* называют математическое ожидание случайной величины

$$D(X)=M[x - M(X)]^2$$

- Для непрерывной величины: $D(X) = \int_{-\infty(a)}^{+\infty(b)} [x - M(X)]^2 \varphi(x) dx$ (6)
- Для дискретной величины: $D(X) = \sum_{i=1}^{n} [x_i - M(x)]^2 P_i.$ (7)
- ***Среднее квадратическое отклонение (СКО) σ(X)***

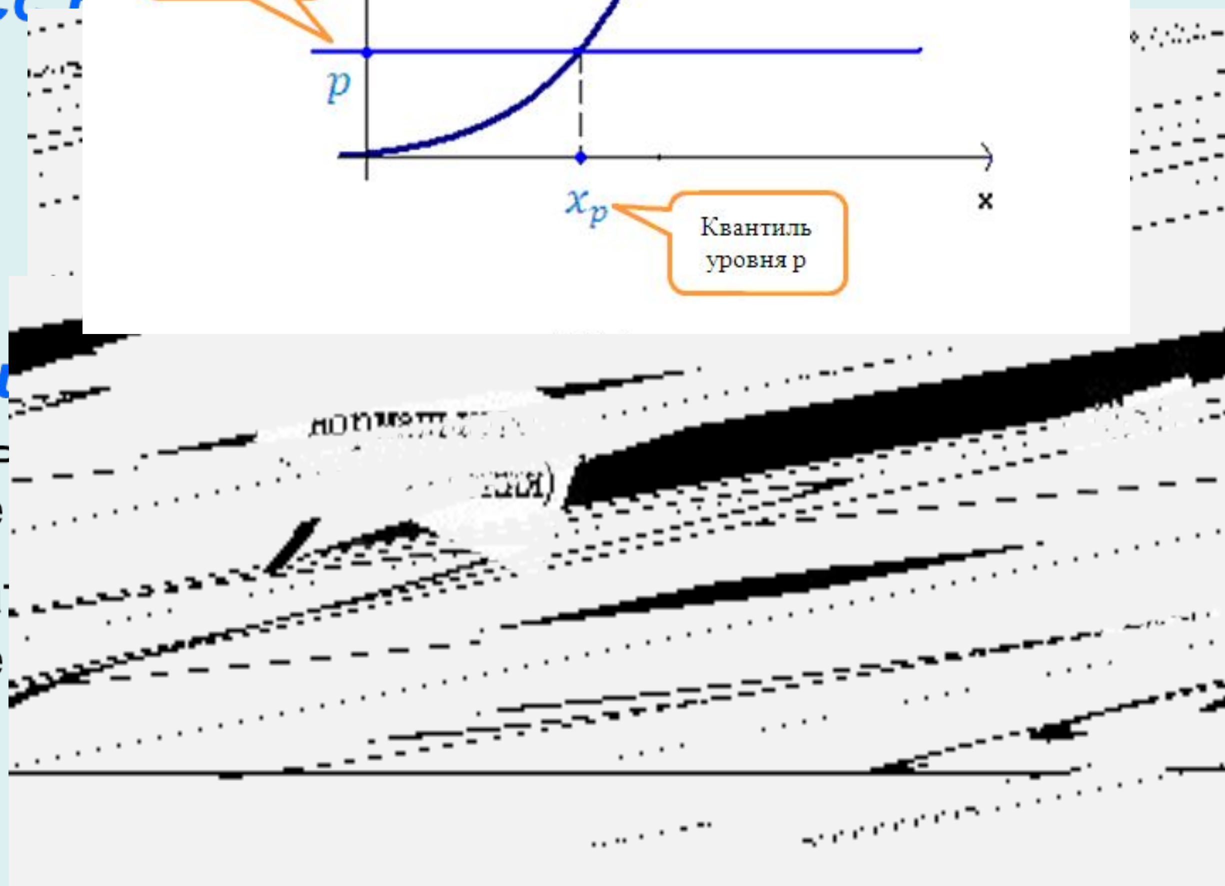
$$\sigma(X) = \sqrt{D(X)} \qquad (8)$$

Важнейш … еделения

- ***Коэффиц***

(9)

- ***Эксцесс***



(10)

- ***Кванти*** … ой величин … ывают решение
- Следоват … димо найти решение

Оценивание генеральных параметров
(*свертывание цифровой информации*)

Требования предъявляемые к оценке параметров распределения (Несмещеность, состоятельность и эффективность):

Состоятельная оценка – это оценка, которая при увеличении числа наблюдений стремится к истинному значению оцениваемого параметра.

Несмещенная оценка – оценка, математическое ожидание которой равно истинному значению оцениваемого параметра.

Эффективная оценка – оценка, имеющая наименьшую дисперсию по сравнению с любой другой оценкой данного параметра.

Оценивание генеральных параметров
(*свертывание цифровой информации*)

Оценки математического ожидания:

<u>*Среднее арифметическое*</u> $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (11)$$

$$\lim_{n \to \infty} \bar{x} = M(X)$$

<u>*Среднее геометрическое*</u> $$\boldsymbol{x_{г}} = \sqrt[n]{\boldsymbol{x_1} \cdot \boldsymbol{x_2} \cdot \cdots \cdot \boldsymbol{x_n}} \qquad (12)$$

<u>*Медиана*</u> $(\tilde{x})$ $\quad x_1 \le x_2 \le x_3 \le \ldots . \le x_n.$

<u>*Мода*</u>

Оценивание генеральных параметров
(*свертывание цифровой информации*)

Оценки генеральной дисперсии:

<u>*Выборочная дисперсия*</u> $S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}, \quad f = n - 1$ (13)

<u>*Стандартное отклонение*</u> $S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$ (14)

<u>*Стандартное отклонение среднего арифметического*</u>

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

$\lim\limits_{n\to\infty} S^2 = D(X)$ и $\lim\limits_{n\to\infty} S = \sigma(X)$

Другие статистические показатели выборок

<u>*Размах*</u> $R = x_{max} - x_{min}$ (15)

<u>*Асимметрия*</u> $$A = \frac{1}{nS^3}\sum_{i=1}^{n}(x_i - \bar{x})^3$$ (16)

<u>*Эксцесс*</u> (17)

$$E = \frac{1}{nS^4}\sum_{i=1}^{n}(x_i - \bar{x})^4 - 3$$

Другие статистические показатели выборок

<u>*Доверительный интервал*</u>

Симметричный интервал с границами ± Δx(P) называется доверительным интервалом случайной погрешности с доверительной вероятностью P, если площадь кривой распределения между абсциссами –Δx и +Δx составляет P-ю часть всей площади под кривой плотности распределения вероятностей.

Другие статистические показатели выборок

__Доверительный интервал погрешности:__

$$\Delta x = \frac{t_{\alpha,f} S(x)}{\sqrt{n}},$$

где $t_{\alpha,f}$ – нормированный показатель распределения, зависящий от числа степеней свободы (n–1) и выбранной доверительной вероятности P.

Он определяется с помощью таблицы α-процентных точек распределения, которая имеет два параметра: $f = n - 1$ и $\alpha = 1 - P$;

__Доверительный интервал истинного (действительного) значения :__

$$\left[\bar{x} - \frac{t_{\alpha,f} S(x)}{\sqrt{n}} ; \bar{x} + \frac{t_{\alpha,f} S(x)}{\sqrt{n}}\right]$$

Интервал, который с определенной вероятностью накрывает истинное значение величины, называется ***доверительным интервалом***, а соответствующая этому интервалу вероятность – ***доверительной вероятностью*** **(P)**.

Теоретические распределения

ДИСКРЕТНЫЕ:

- Биномиальное распределение
- Распределение Пуассона

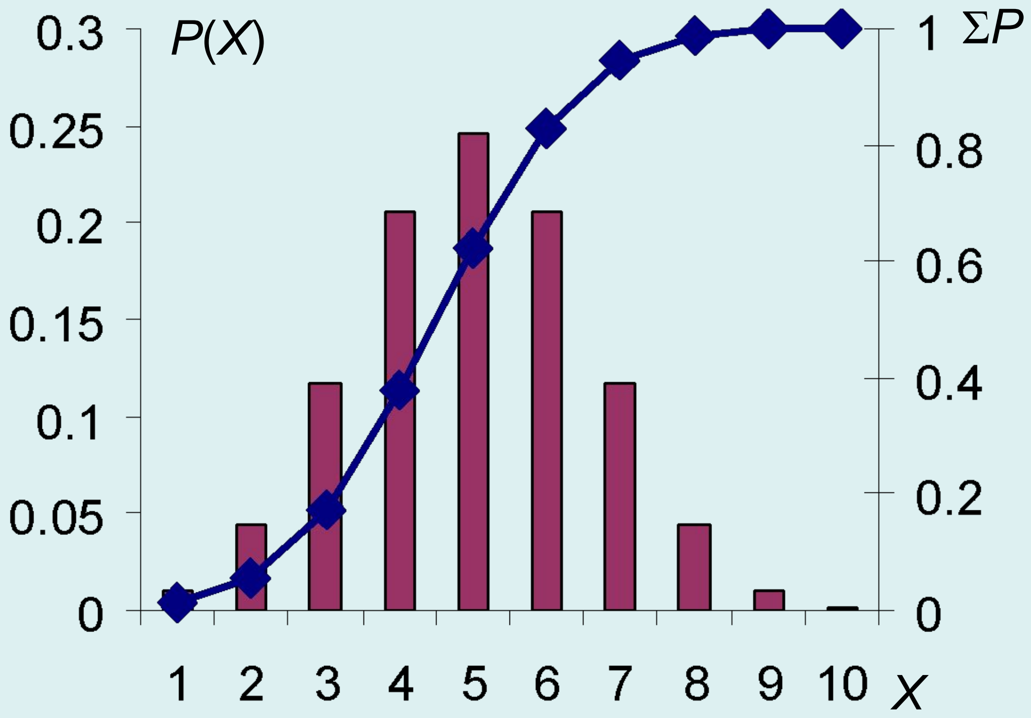
НЕПРЕРЫВНЫЕ:

- Нормальное
- Распределение Стьюдента
- Логарифмически нормальное
- Распределение Пирсона
- Распределение Фишера

Знание распределений необходимо для построения статистических выводов из выборочных данных: оценивание, проверка гипотез, отыскание допустимые отклонений, вероятностей ошибок и др.

Биномиальное распределение

$$P_n(k) = C_n^k p^k (1-p)^{n-k}$$



Параметры распределения
$M(X) = np;$
$D(X) = np(1-p)$

Вычисляют вероятность того, что при *n* испытаниях успешными окажутся *k* испытаний, при условии, что вероятность единичного успеха равна *p*.

Используют для вычисления погрешности при пробоотборе штучных образцов:

$$\sigma_r = \frac{\sigma}{M(X)} = \sqrt{\frac{1-p}{np}}$$

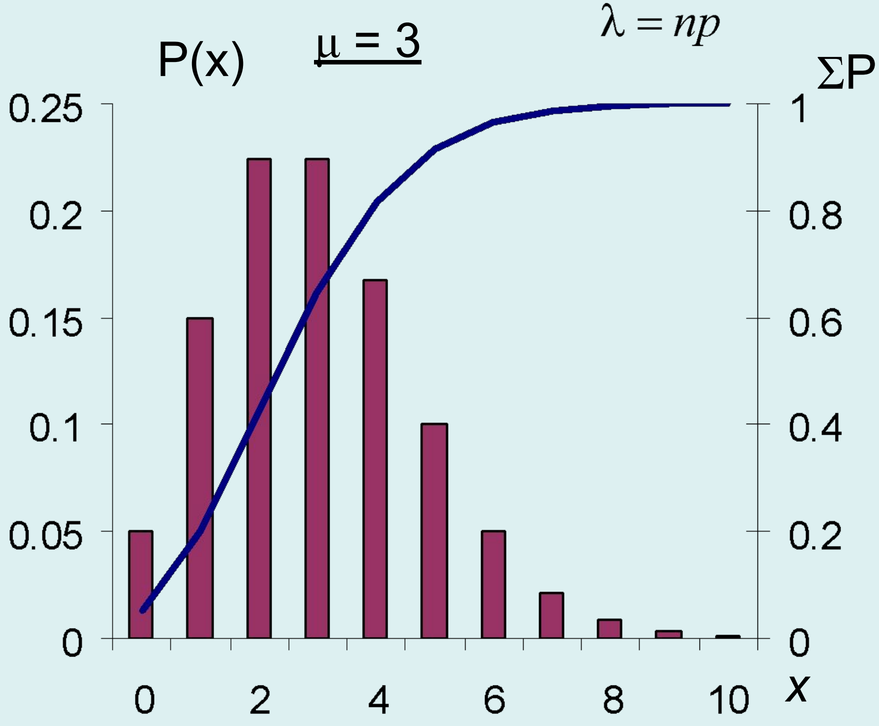
или объема выборки при заданной погрешности пробоотбора

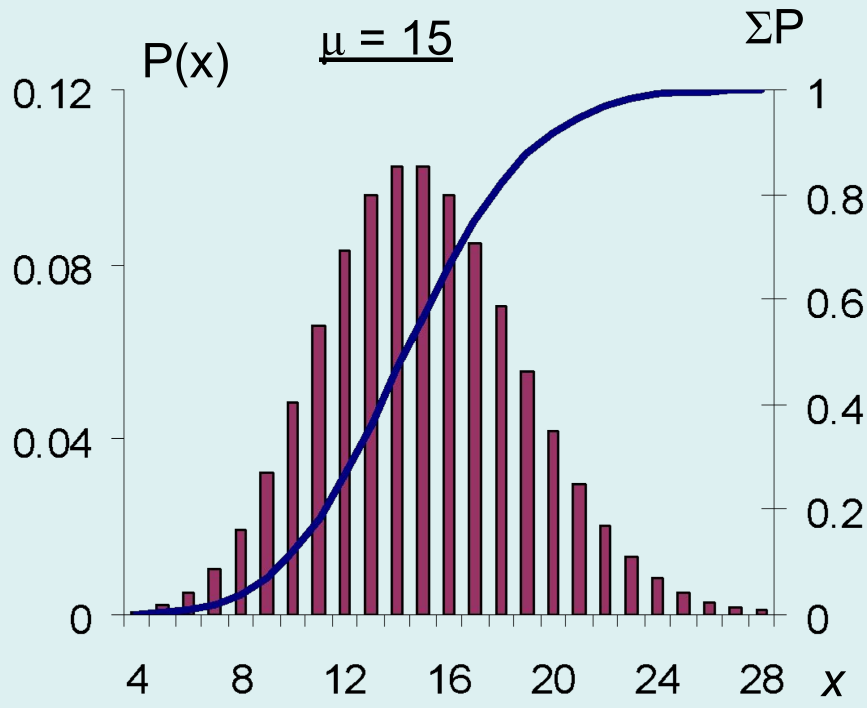
$$n = \frac{\sigma}{M(X)} = \frac{1-p}{\sigma_r^2 p}$$

Распределение Пуассона

$$P_n(k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad k - \text{число появления событий в n незвисимых испытаниях;}$$

$$\lambda = np$$





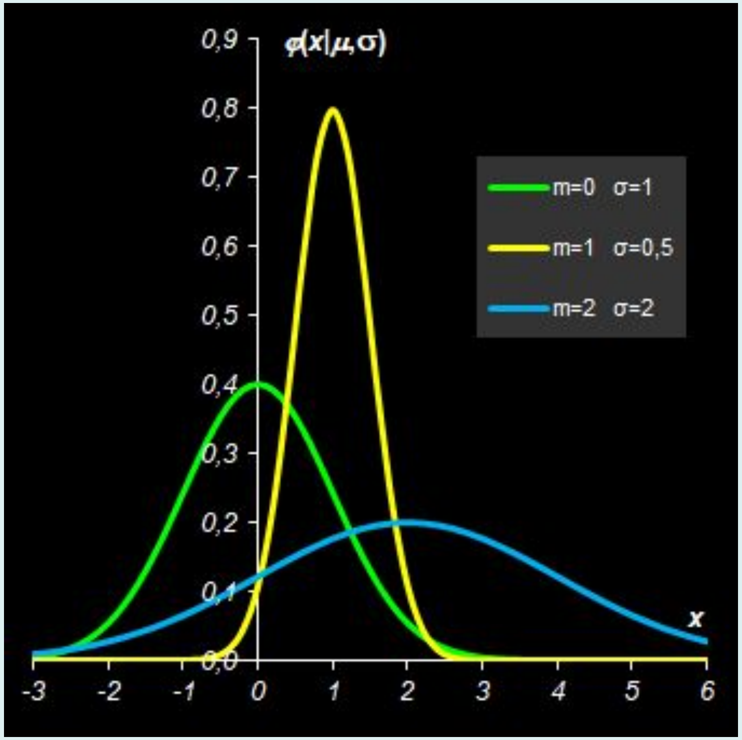
Параметры распределения
$M(X) = \lambda;$
$D(X) = \lambda \qquad\qquad \lambda = np$

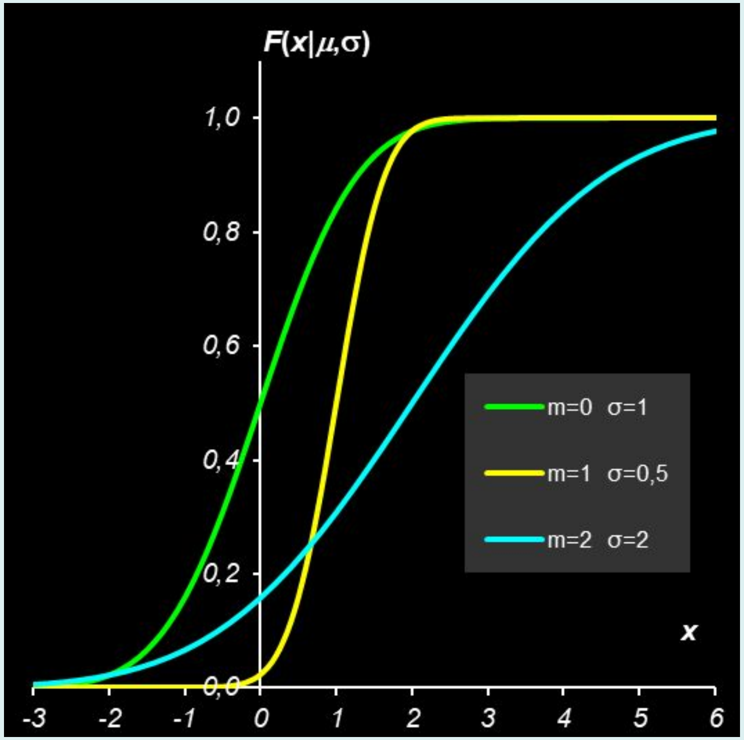
Применяют при регистрации скоростей счета частиц в радиохимии, числа квантов в рентгеноспектральном анализа, числа структурных элементов шлифов при фазовом анализа и др.

Нормальное распределение

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(z-\mu)^2}{2\sigma^2} \right) \qquad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2} \right) dt$$

μ – любое действительное число и $\sigma > 0$





Параметры распределения
$M(X) = \mu;$
$D(X) = \sigma^2$

Нормальное распределение

Основные пр

1) $\varphi(x) \geq 0$;

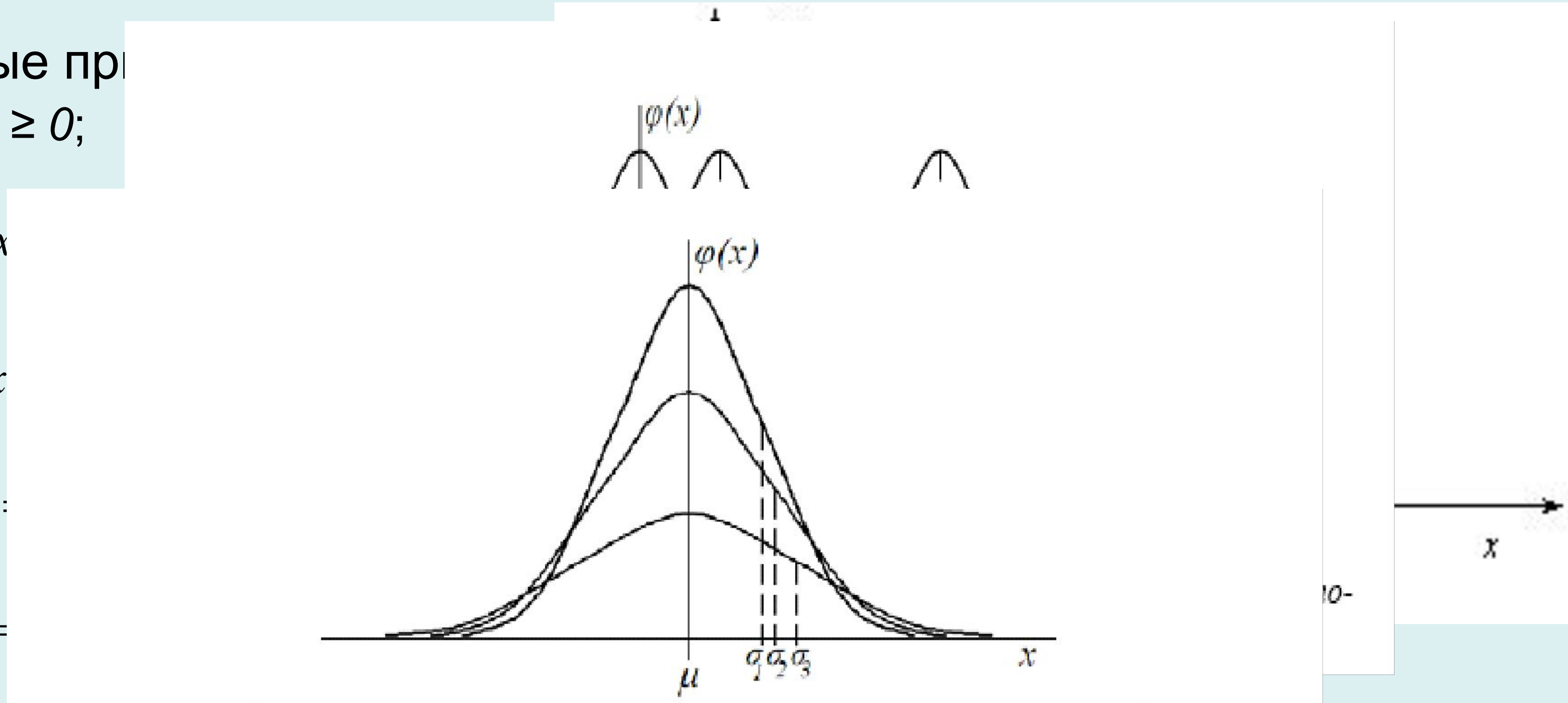
2) $\int_a^b \varphi(x$

3) $\int_a^{\mu+x} \varphi(x$

4) $\frac{d\varphi(x)}{dx}$

5) $\frac{d^2\varphi(x)}{dx^2}$

6) μ (σ=



Функции плотности вероятности нормального распределения при постоянной μ, σ₁< σ₂< σ₃.

7. σ определяет степень «размытости» кривой.
8. При $\mu = 0$, $\sigma = 1$ (*Стандартное нормальное распределение*)

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

Станда... ...деление

- Коэффициент

Φ

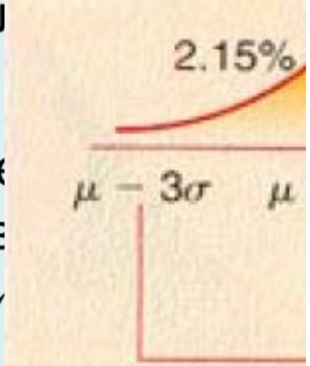
φ

$= -\Phi(u)$

Дл

ве

Дл



Ве

се

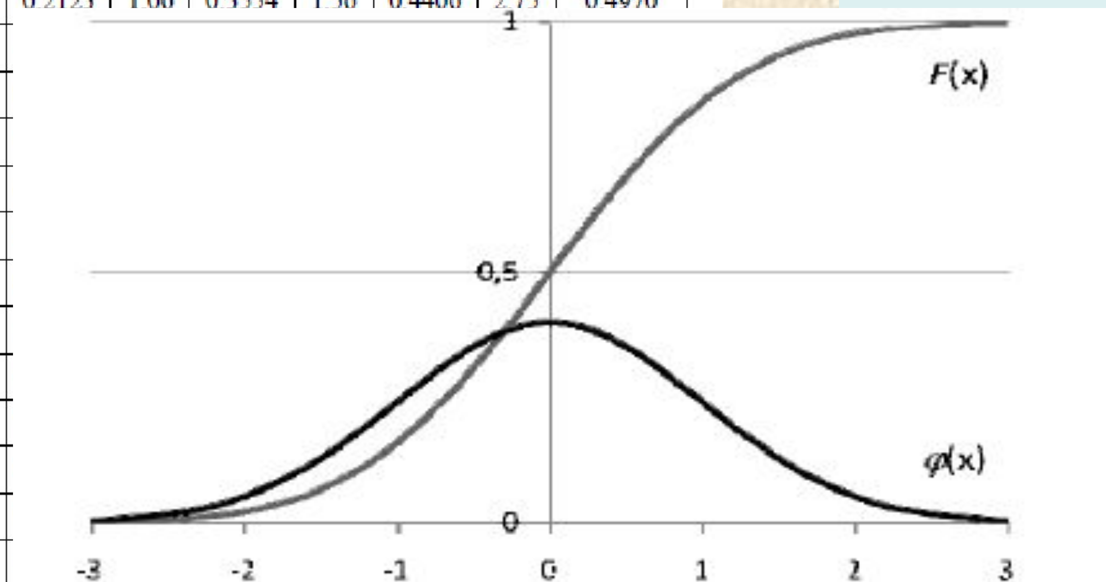
P(

гд

Пр

Приложение 1. *Значения функции Лапласа* $\Phi(u) = \frac{1}{\sqrt{2\pi}}\int_0^u e^{\frac{-u^2}{2}}\,du.$

u	Φ(u)	u	Φ(u)	u	Φ(u)	u	Φ(u)	u	Φ(u)
0,00	0,0000	0,42	0,1628	0,92	0,3212	1,42	0,4222	2,40	0,4918
0,01	0,0040	0,44	0,1700	0,94	0,3264	1,44	0,4251	2,45	0,4929
0,02	0,0080	0,46	0,1772	0,96	0,3315	1,46	0,4279	2,50	0,4938
0,03	0,0120	0,48	0,1844	0,98	0,3365	1,48	0,4306	2,55	0,4946
0,04	0,0160	0,50	0,1915	1,00	0,3413	1,50	0,4232	2,60	0,4953
0,05	0,0199	0,52	0,1985	1,02	0,3461	1,52	0,4357	2,65	0,4959
0,07	0,0279	0,54	0,2054	1,04	0,3508	1,54	0,4382	2,70	0,4965
0,08	0,0319	0,56	0,2123	1,06	0,3554	1,56	0,4406	2,75	0,4970
0,09	0,0359	0,58							
0,10	0,0398	0,60							
0,12	0,0478	0,62							
0,14	0,0557	0,64							
0,16	0,0636	0,66							
0,18	0,0714	0,68							
0,20	0,0793	0,70							
0,22	0,0871	0,72							
0,24	0,0948	0,74							
0,26	0,1026	0,76							
0,28	0,1103	0,78							
0,30	0,1179	0,80							
0,32	0,1255	0,82	0,2939	1,32	0,4066	2,15	0,4843	4,00	0,499968
0,34	0,1331	0,84	0,2995	1,34	0,4099	2,20	0,4861	4,20	0,499987
0,36	0,1406	0,86	0,3051	1,36	0,4131	2,25	0,4878	4,40	0,499995
0,38	0,1480	0,88	0,3106	1,38	0,4162	2,30	0,4893	4,50	0,499997
0,40	0,1554	0,90	0,3159	1,40	0,4192	2,35	0,4906	5,00	0,499997



Стандартное нормальное распределение

Для среднего арифметического результатов измерений

$$\sigma(\bar{x}) = \sqrt{D(\bar{x})} = \sqrt{\frac{D(x)}{n}} = \frac{\sigma(x)}{\sqrt{n}}$$

Справедливо:

$$P = \{\mu - u\sigma/\sqrt{n} \leq \bar{x} \leq \mu + u\sigma/\sqrt{n}\} = 2\Phi(u)$$

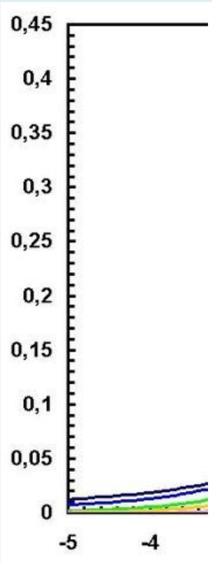
Для оценки математического ожидания μ случайной величины X, распределенной по нормальному закону, при известном среднем квадратическом отклонении σ служит *доверительный интервал*

$$\bar{x} - u\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + u\frac{\sigma}{\sqrt{n}}$$

где $u\frac{\sigma}{\sqrt{n}} = \delta$ – точность оценки,

n – объем выборки,
$\bar{x}$– выборочное среднее,
u – аргумент функции Лапласа, при котором $\Phi(u) = (1 - P)/2$

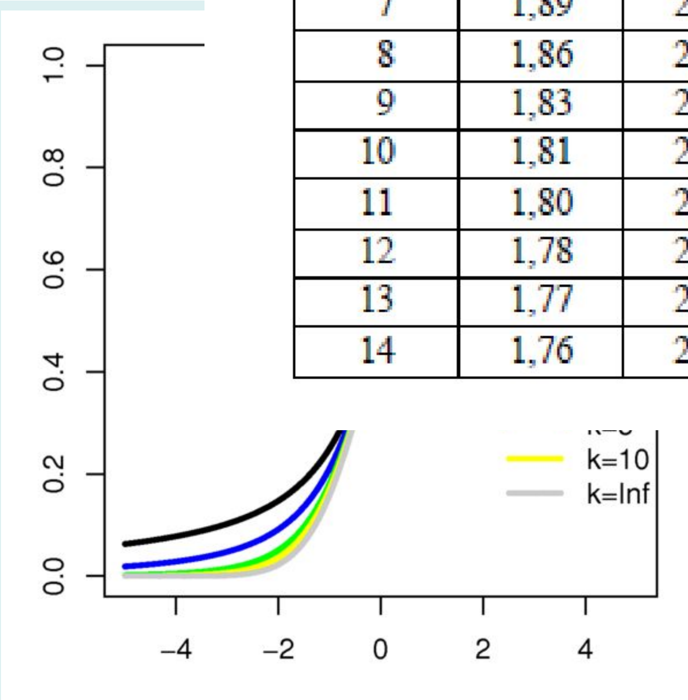
Распределения Стьюдента



Приложение 2. *Коэффициенты распределения Стьюдента t при различных степенях свободы f.*

f	t			f	t		
	$P = 0{,}90$	$P = 0{,}95$	$P = 0{,}99$		$P = 0{,}90$	$P = 0{,}95$	$P = 0{,}99$
1	6,31	12,7	63,9	15	1,75	2,13	2,95
2	2,92	4,30	9,93	16	1,75	2,12	2,92
3	2,35	3,18	5,84	17	1,74	2,11	2,90
4	2,13	2,78	4,60	18	1,73	2,10	2,88
5	2,02	2,57	4,03	19	1,73	2,09	2,86
6	1,94	2,45	3,71	20	1,72	2,09	2,85
7	1,89	2,37	3,50	21	1,72	2,08	2,83
8	1,86	2,31	3,36	22	1,72	2,07	2,82
9	1,83	2,26	3,25	25	1,71	2,06	2,79
10	1,81	2,23	3,17	30	1,70	2,04	2,75
11	1,80	2,20	3,11	40	1,68	2,02	2,70
12	1,78	2,18	3,06	60	1,67	2,00	2,66
13	1,77	2,16	3,01	120	1,66	1,98	2,62
14	1,76	2,14	2,98	∞	1,6479	1,9647	2,5786

$$\frac{\ldots\mu)\sqrt{n}}{\sigma}$$

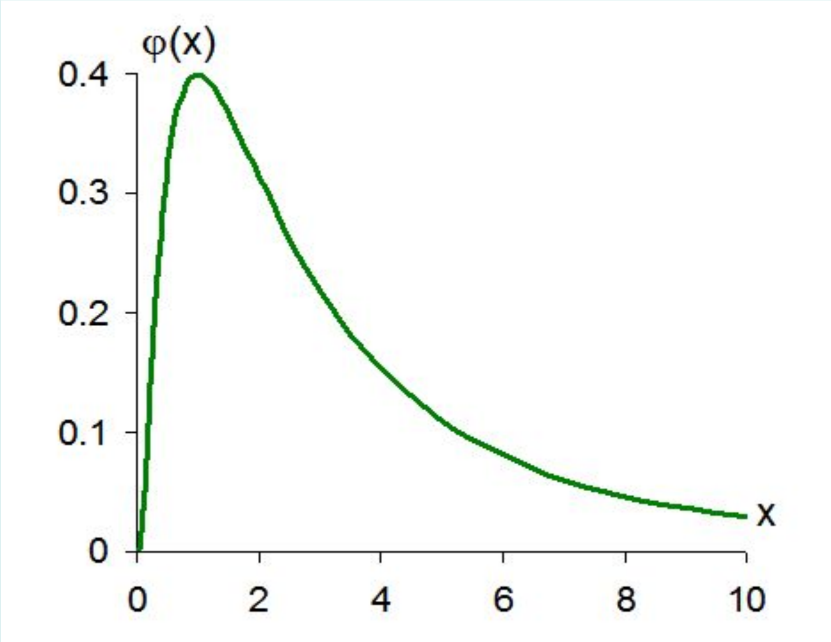


ия

$M(X) = 0$ при $n > 1$

$$D(X) = \frac{n}{n-2} \quad \text{при } n > 2$$

Логнормальное распределение



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

Логнормальное распределение имеет место в следующих случаях:

- в инструментальных методах анализа, в которых аналитический сигнал пропорционален логарифму концентрации определяемого компонента;
- при определении очень низких концентраций аналитов;
- при очень большом разбросе результатов;
- при измерении времени.

$$x_{\text{г}} = \sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n}.$$

$$S_{lgx} = \sqrt{\frac{\sum_{i=1}^{n}(lgx_i - \overline{lgx})^2}{n-1}}$$

$$\overline{lgx} - t_{(P,f)}S_{lgx} \le lg\mu \le \overline{lgx} + t_{(P,f)}S_{lgx}$$

$$\frac{x}{10^{t_{(P,f)}S_{lgx/\sqrt{n}}}} \le \mu \le x \cdot 10^{t_{(P,f)}S_{lgx/\sqrt{n}}}$$

Распределение χ^2 (Пирсона)

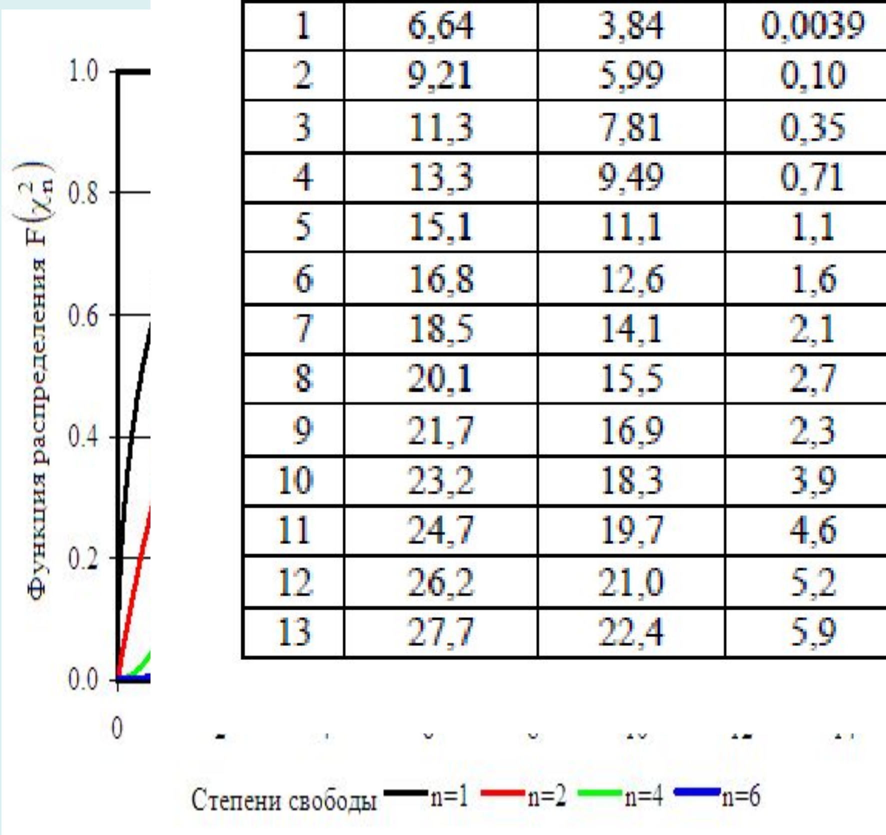
Случа…

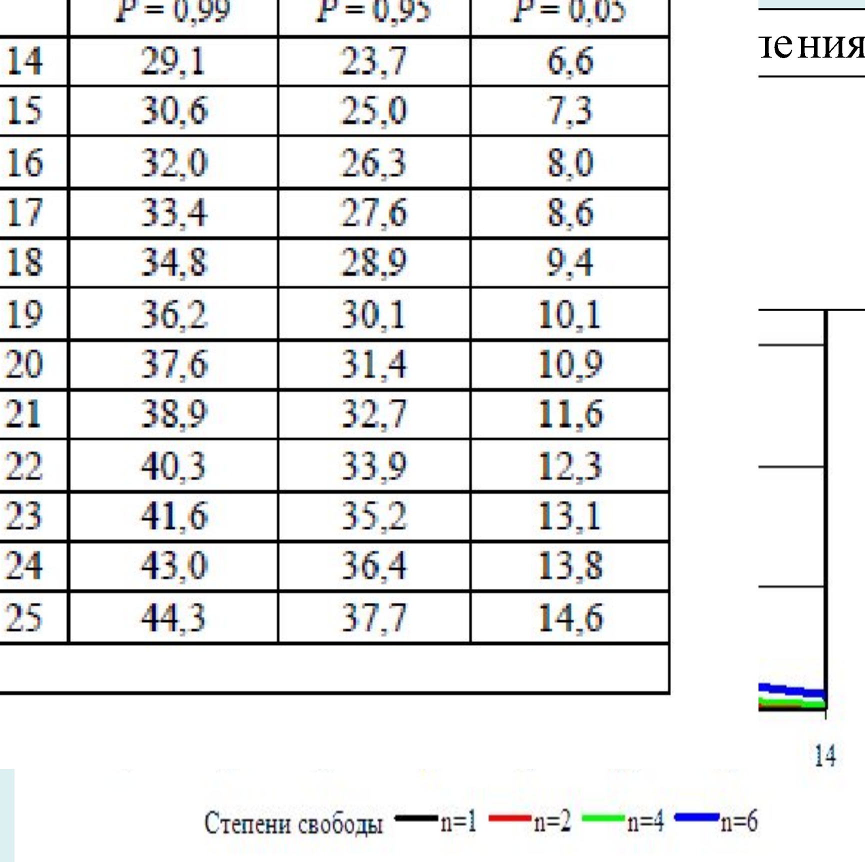
распре…

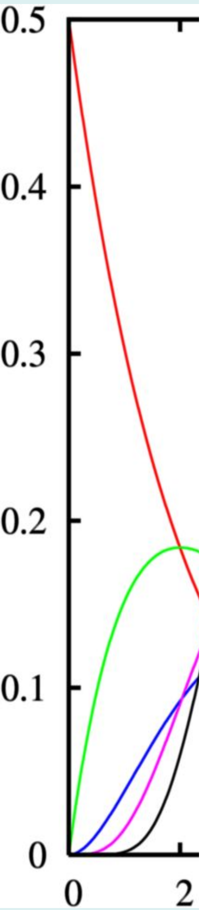
$\chi^2(f) = \ldots$ … $n-1$

Приложение 3. *Значения критерия Пирсона $\chi^2_{\text{кр}}$ для различных степеней свободы f и доверительной вероятности P.*

f	$\chi^2_{\text{кр}}$			f	$\chi^2_{\text{кр}}$		
	$P=0{,}99$	$P=0{,}95$	$P=0{,}05$		$P=0{,}99$	$P=0{,}95$	$P=0{,}05$
1	6,64	3,84	0,0039	14	29,1	23,7	6,6
2	9,21	5,99	0,10	15	30,6	25,0	7,3
3	11,3	7,81	0,35	16	32,0	26,3	8,0
4	13,3	9,49	0,71	17	33,4	27,6	8,6
5	15,1	11,1	1,1	18	34,8	28,9	9,4
6	16,8	12,6	1,6	19	36,2	30,1	10,1
7	18,5	14,1	2,1	20	37,6	31,4	10,9
8	20,1	15,5	2,7	21	38,9	32,7	11,6
9	21,7	16,9	2,3	22	40,3	33,9	12,3
10	23,2	18,3	3,9	23	41,6	35,2	13,1
11	24,7	19,7	4,6	24	43,0	36,4	13,8
12	26,2	21,0	5,2	25	44,3	37,7	14,6
13	27,7	22,4	5,9				







Значения критерия Фишера (F-критерия) для уровня значимости q = 5%

ν_1 – число степеней свободы большей дисперсии; ν_2 – число степеней свободы меньшей дисперсии

ν_2	ν_1											
	1	2	3	4	5	6	7	8	9	10	15	20
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	245,9	248,0
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,43	19,45
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,70	8,66
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,86	5,80
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,62	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,15
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,94
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,77
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,72	2,65
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,62	2,54
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,53	2,46
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,46	2,39
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,33
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,35	2,28
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,31	2,23
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,27	2,19
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,23	2,16
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,20	2,12
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,18	2,10
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,15	2,07
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,13	2,05
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,11	2,03
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,09	2,01
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,07	1,99
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,06	1,97
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,04	1,96
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,03	1,94
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,01	1,93

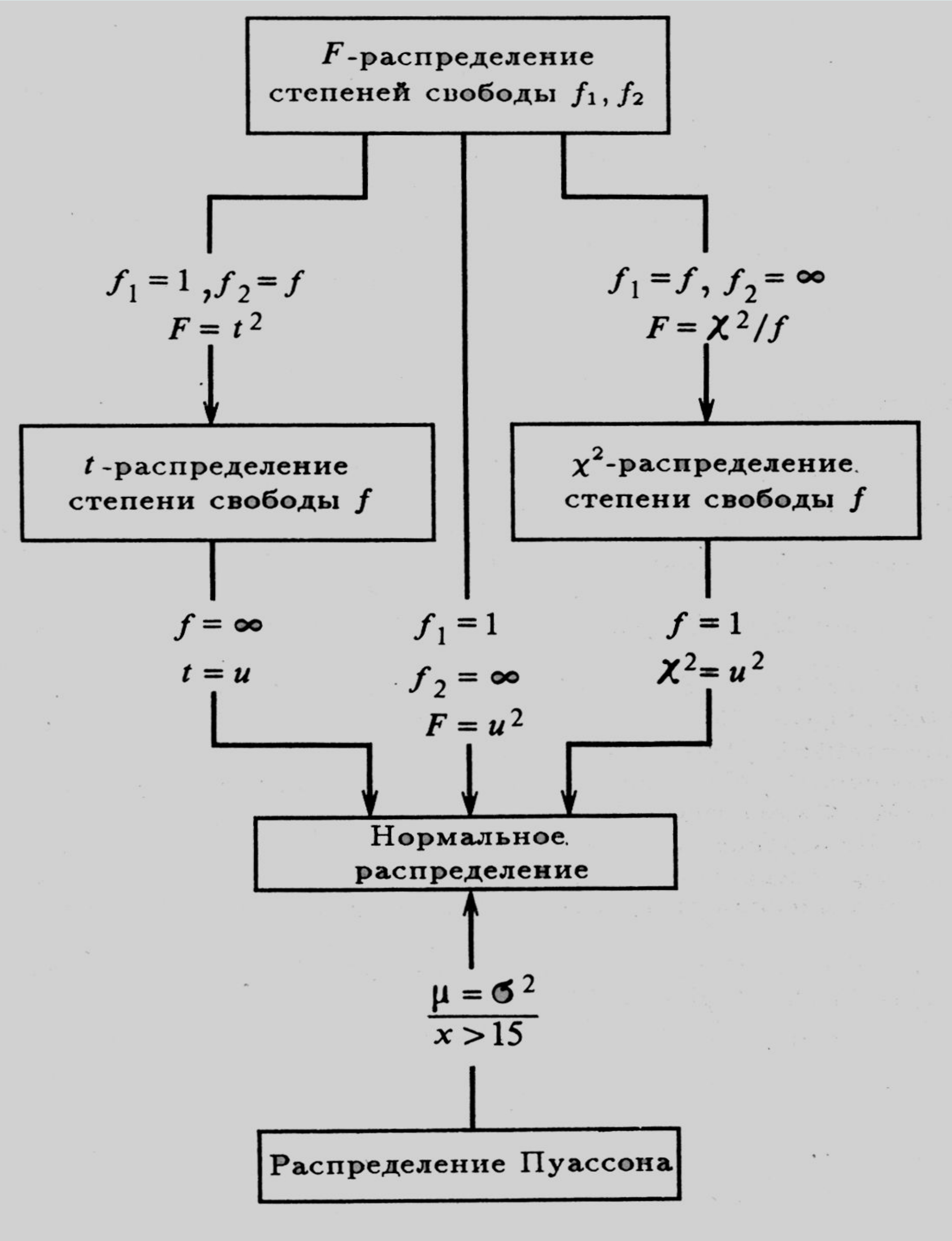
ия

2

$f_2)$

$_2 - 3)$

Связь между отдельными теоретическими распределениями



Результат измерения и оценка его случайной погрешности

- Однократные измерения

Оценка доверительной вероятности того, что единичный результат анализа отклоняется от своего математического ожидания не больше, чем на заданную величину $\Delta x = \pm z\sigma$, в соответствии с неравенством Чебышева имеет следующий вид:

$$P\{\bar{x} - z\sigma_x \le x_i \le \bar{x} + z\sigma_x\} \ge 1 - \frac{1}{z^2}$$

$$[x_i - 1{,}96\,{}^{\alpha}S_x;\ x_i + 1{,}96\,{}^{\alpha}S_x];\ \varepsilon = \pm\, 1{,}96\,{}^{\alpha}S_x\ (P = 0{,}95).$$

При неизвестном виде распределения измеряемой величины согласно неравенству Чебышева интервалы и случайные погрешности равны

$$[x_i - 4{,}5\,{}^{\alpha}S_x;\ x_i + 4{,}5\,{}^{\alpha}S_x];\ \varepsilon = \pm\, 4{,}5\,{}^{\alpha}S_x\ (P = 0{,}95)$$

Результат измерения и оценка его случайной погрешности

Многократные измерения

Вид распределения	Результат анализа	Случайная погрешность (P= 0,95)
Нормальное (n > 20)		
Нормальное (Стьюдента) (2 > n > 20)		
Усеченное нормальное		
Логнормальное		
Неизвестное		

Разделы темы для СРС

1. Усеченное нормальное распределение.
2. Равномерное распределение.
3. Треугольное распределение.
4. Генерирование случайных чисел