М.А. Авдыев

Теория вероятности и математическая статистика

Почти просто!

marat@emediator.ru

#ФизикаДляМенеджеров

Союз СЦМ 2023

Две конкурирующие газеты в городе - миллионнике

Данные Росстат: пол, возраст, доход, род занятий / профессия . . .

Вероятно *Да*, вероятно *Нет*

p + q = 1

Орёл - p

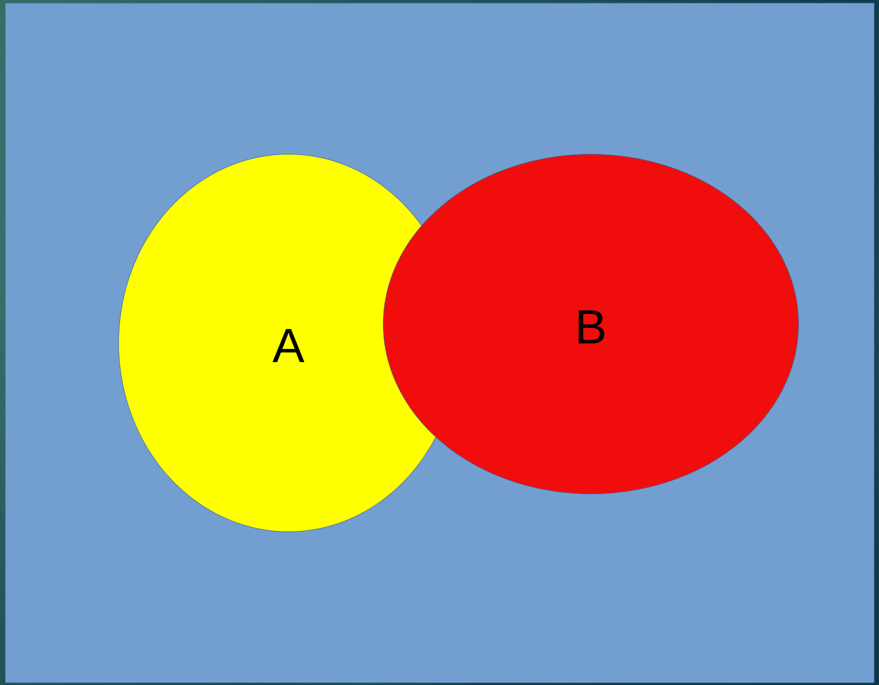
Решка - q

$$P\{A\} = \lim_{N \to \infty} \frac{N_A}{N}$$

Свойства вероятностей

$$0 \le P\{A\} \le 1$$

$$P\{A\} + P\{\bar{A}\} = 1$$



$$P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\}$$

Условная вероятность

$$P\{A|B\} = \lim_{N\to\infty} \frac{N_{AB}}{N} \ldots \frac{N_B}{N} \frac{N_{AB}}{N_B}$$

$$P\{A\&B\} = P\{A|B\}P\{B\} = P\{B|A\}P\{A\}$$

$p_1 = 0{,}2$, $p_2 = 0{,}3$ $p_3 = 0{,}5$

$q_1 = 0{,}8$, $q_2 = 0{,}7$ $q_3 = 0{,}5$

$p = \ (1 — q_1q_2q_3) = 0{,}72$

$$P\{A\&B\} = P\{A\}P\{B\}$$

A и B независимы!

Игра в *снэп.* 2 колоды карт по 52.

p (нет совпадения)?

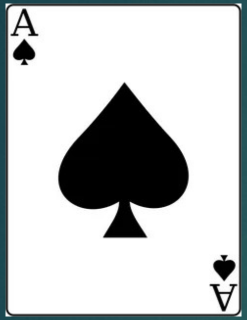
$$(p+q)^n = \sum_{k=1}^{k=n} C_n^k p^k q^{n-k}$$

$$q = 1 - p = 1 - \frac{1}{n}$$

$$(1-\frac{1}{n})^n = \lim(1+x)^{-\frac{1}{x}} = \lim\frac{1}{(1+x)^{\frac{1}{x}}} = \frac{1}{e}$$

p = 0,3679

Вероятность (не)зависимых

P? 

$$P(A) = \frac{1}{4}$$

$$P(AB) = \frac{1}{52}$$

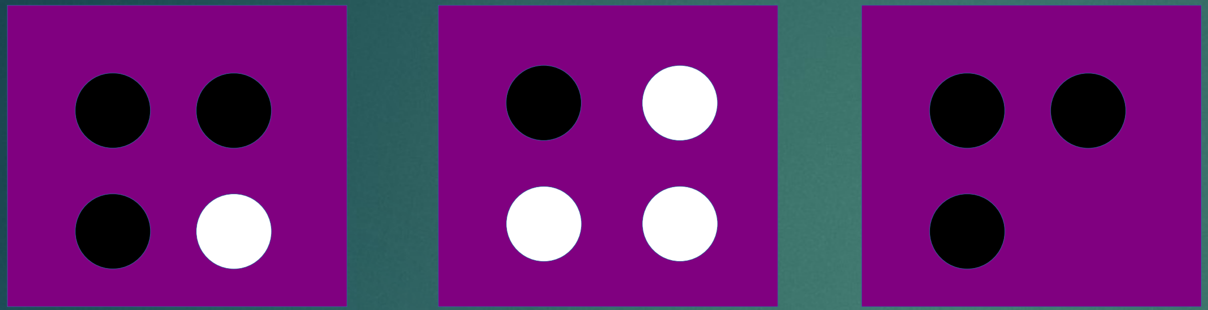
$$P(B) = \frac{1}{13}$$

$$P(A) = \frac{4}{53} \qquad P(B) = \frac{13}{53} \qquad P(AB) = \frac{1}{53}$$

$$P(A) * P(B) = \frac{4}{53} * \frac{13}{53} = \frac{1}{53} - (\frac{1}{53})^2$$

Формула полной вероятности

$$P\{black\} = \frac{1}{3} * \frac{3}{4} + \frac{1}{3} * \frac{1}{4} + \frac{1}{3} * \frac{1}{1} = \frac{2}{3}$$



$$P\{A\} = P\{A|H_1\}P\{H_1\} + P\{A|H_2\}P\{H_2\} + P\{A|H_3\}P\{H_3\}$$

Формула Байеса. Контроль качества продукции.

$$P\{A\}P\{H_i|A\} = P\{H_i\}P\{A|H_i\}$$

$\frac{0{,}9}{800}$

$\frac{0{,}5}{600}$

$\frac{0{,}2}{900}$

$$P\{A\} = P\{A|H_1\}P\{H_1\} + P\{A|H_2\}P\{H_2\} + P\{A|H_3\}P\{H_3\}$$

$$P\{H_i|A\} = \frac{P\{A|H_i\}P\{H_i\}}{P\{A|H_1\}P\{H_1\} + P\{A|H_2\}P\{H_2\} + \ldots + P\{A|H_k\}P\{H_3\}}$$

Контроль качества продукции. Численный пример

0,9

800

0,5

600

0,2

900

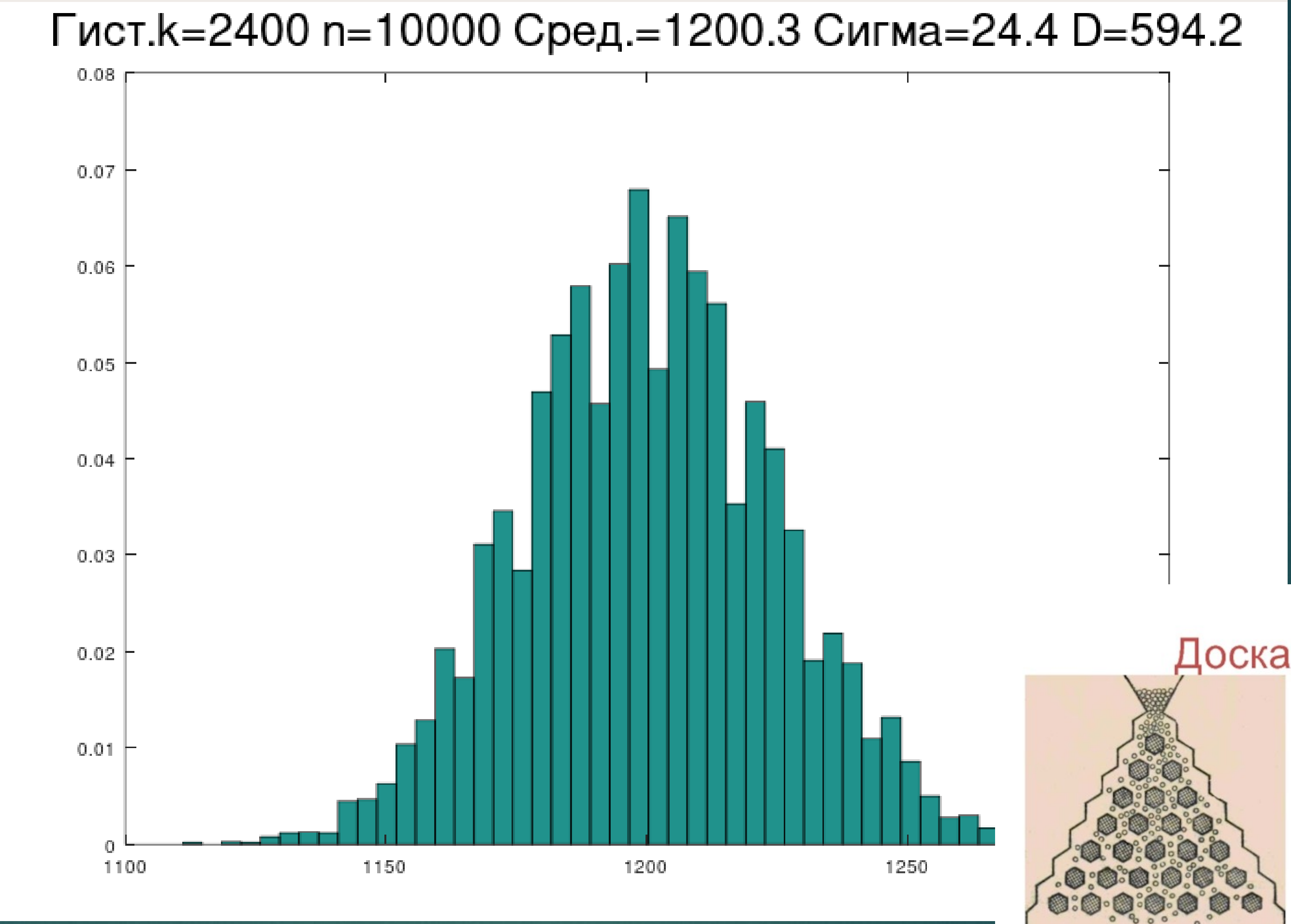
$$P\{H_i|A\} = \frac{P\{A|H_i\}P\{H_i\}}{P\{A|H_1\}P\{H_1\} + P\{A|H_2\}P\{H_2\} + \ldots + P\{A|H_k\}P\{H_3\}}$$

$$\frac{0{,}2 \cdot 900/2300}{0{,}9 \cdot 800/2300 + 0{,}5 \cdot 600/2300 + 0{,}2 \cdot 900/2300} = 0{,}15.$$

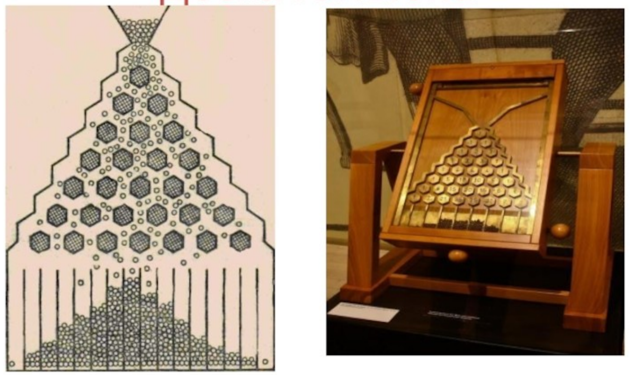
Биноминальное распределение

$$(p+q)^n = \sum_{k=1}^{k=n} C_n^k p^k q^{n-k}$$

$$C_n^k = \frac{k!}{n!(n-k)!}$$



Доска Гальтона



Математическое ожидание и дисперсия

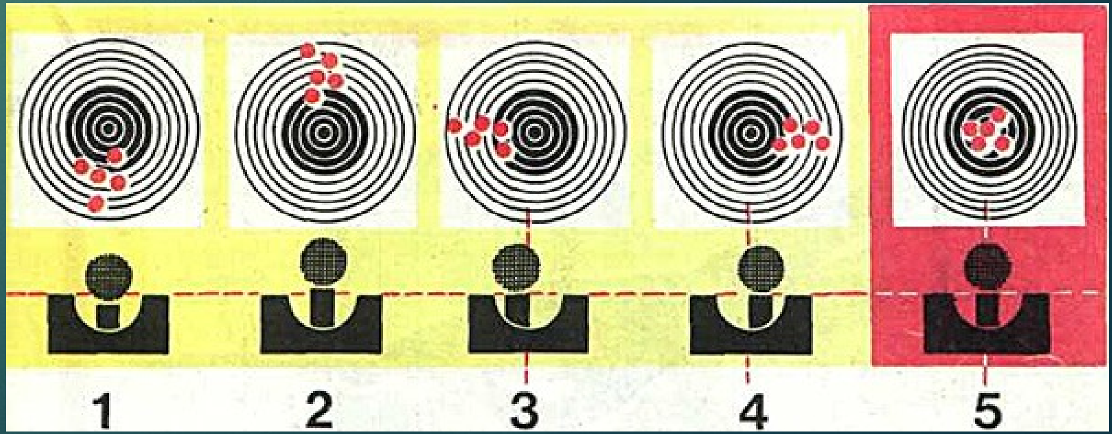
$$M(X) = \sum_{i=1}^{\infty} x_i p_i$$

$M(x) = p_1x_1 + p_2x_2 + \ldots p_nx_n$

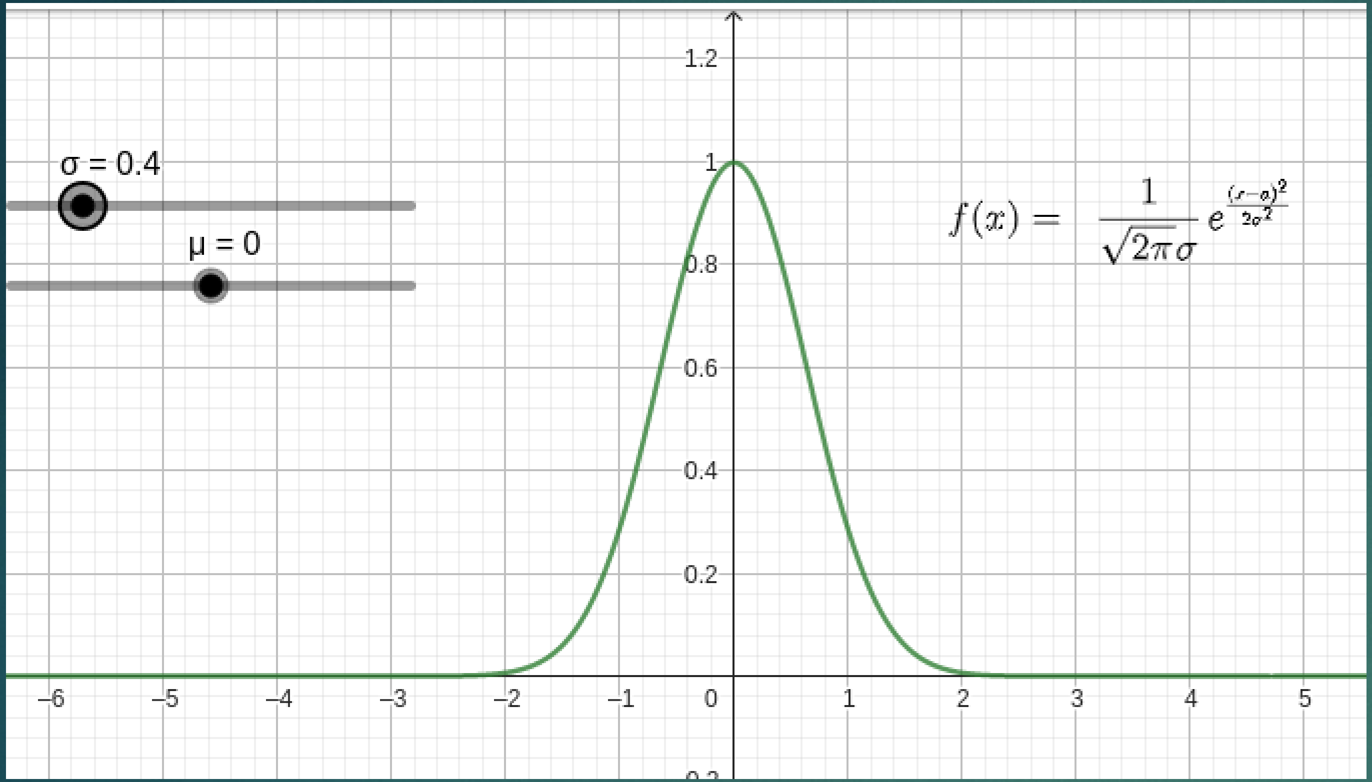
$$D(x) = M\left((x - M(x))^2\right) = M(x^2) - (M(x))^2$$

$D(x + y) = D(x) + D(y)$ где x, y - независимы
$D(x - y) = D(x) + D(y)$
$D(C + x) = D(x)$
$D(Cx) = C^2 * D(x)$

$$\sigma = \sqrt{D}$$



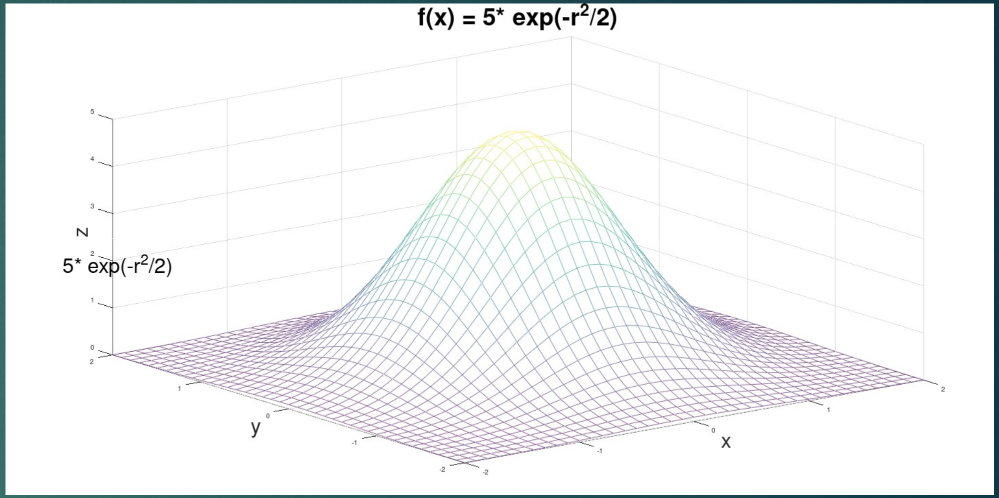
Нормальное распределение



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

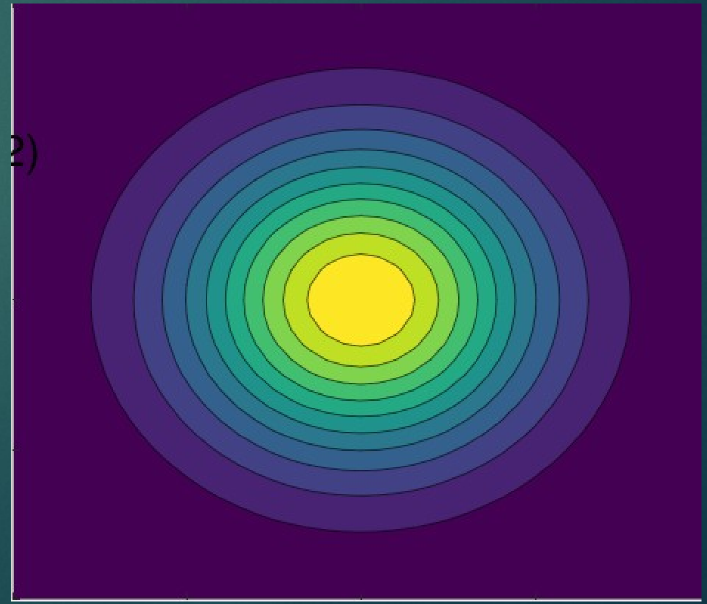
Интеграл вероятности

$$y(x) = \iint e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}} dxdy = 2\pi \int_0^\infty e^{-\frac{r^2}{2}} rdr = 2\pi$$



$$Erf(z) = \sqrt{\frac{2}{\pi}} \int_0^z e^{-\frac{s^2}{2}} ds = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} ds$$

$$t = \sqrt{2}s$$



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

Дисперсия биноминального распределения. Локальная теорема Лапласа

Одно испытание

$D = 1*p - p^2 = p(1-p) = pq$

Серия n испытаний

$D = npq$

Если вероятность p появления события A в каждом испытании постоянна
И отлична от нуля и единицы, то вероятность P (k) того, что событие A появится в N испытаниях ровно k раз, приближенно равна (тем точнее, чем больше N) значению функции

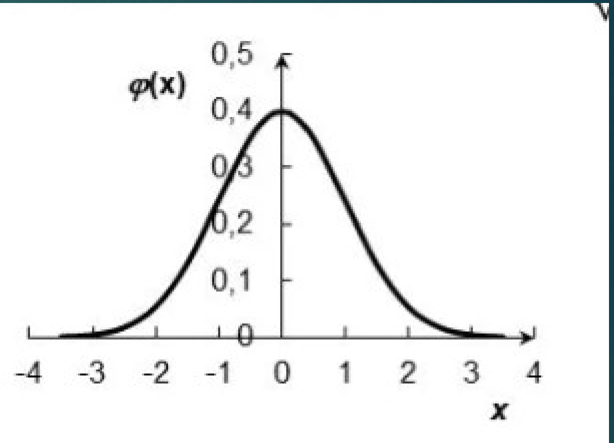
$$y(x) \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Локальная теорема Муавра-Лапласа

вероятность P (k) того что событие A появится в n испытаниях ровно k раз

$$y(x) = \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$x = \frac{k - np}{\sqrt{npq}}$$



Интегральная теорема Муавра-Лапласа

вероятность P (k_1, k_2) того что событие A появится в n испытаниях от k_1 до k_2 раз

$$P(k_1, k_2) \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{z^2}{2}} dz = \Phi(x_2) - \Phi(x_1)$$

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} \qquad x_2 = \frac{k_2 - np}{\sqrt{npq}}$$

Ф(

Пример использования т. Лапласа

Вероятность того, что деталь не прошла проверку ОТК, равна р = 0,2 Найти вероятность того, что среди 400 случайно отобранных деталей окажется непроверенных от 70 до 1 00 деталей.

Р е ш е н и е.

По условию, р = О,2; q = 0,8; n = 400; k1 = 70; k2 = 100. Воспользуемся интегральной теоремой Лапласа:

$$x' = \frac{k_1 - np}{\sqrt{npq}} = \frac{70 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = -1{,}25;$$

$$x'' = \frac{k_2 - np}{\sqrt{npq}} = \frac{100 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = 2{,}5.$$

Таким образом, имеем

$$P_{400}(70,\ 100) = \Phi(2{,}5) - \Phi(-1{,}25) = \Phi(2{,}5) + \Phi(1{,}25).$$

Ф(2,5) = 0,4938   Ф(1,25) = 0,3944

$P_{400}(70, 100) = 0{,}4938 + 0{,}3944 = 0{,}8882 \approx 89\%$

Правило трёх σ. Уровень значим. α

$$P(a-\delta<x<a+\delta)=\frac{1}{\sqrt{2\pi}}\int_{\alpha}^{\beta}e^{-\frac{t^2}{2}}dt=2\Phi(\frac{\delta}{\sigma})dt$$

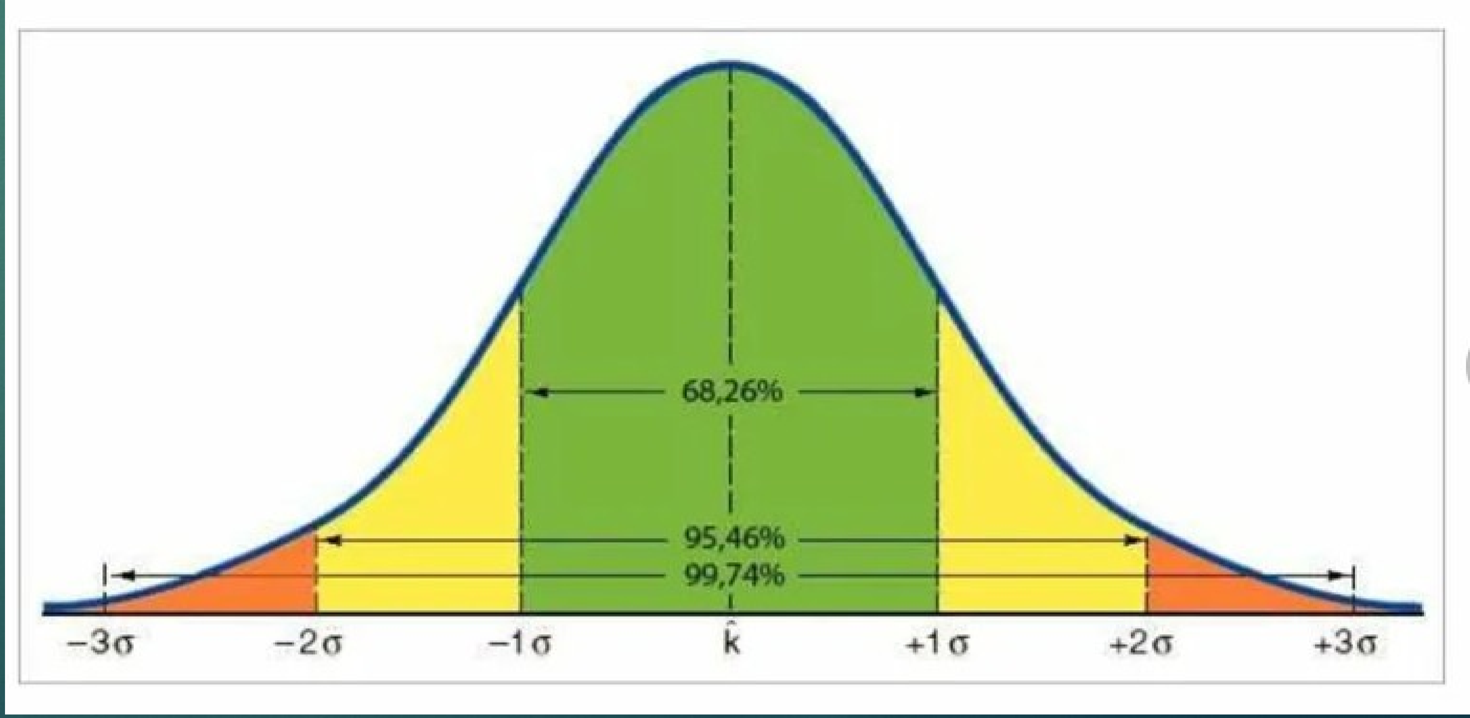
$$t=\frac{x-a}{\sigma}\qquad \alpha=\frac{a-\delta-a}{\sigma}\qquad \beta=\frac{a+\delta-a}{\sigma}$$

N(0,1)

P|x — a| < σ = 2Ф(1) = 0,6826

P|x — a| < 2σ = 2Ф(2) = 0,9544

P|x — a| < 3σ = 2Ф(3) = 0,9972



Распределение Пуассона

$$P(k) = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}p^k(1-p)^{n-k}$$

$$P(k) = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k(1-\frac{\lambda}{n})^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}$$

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

Поток событий интенсивностью λ

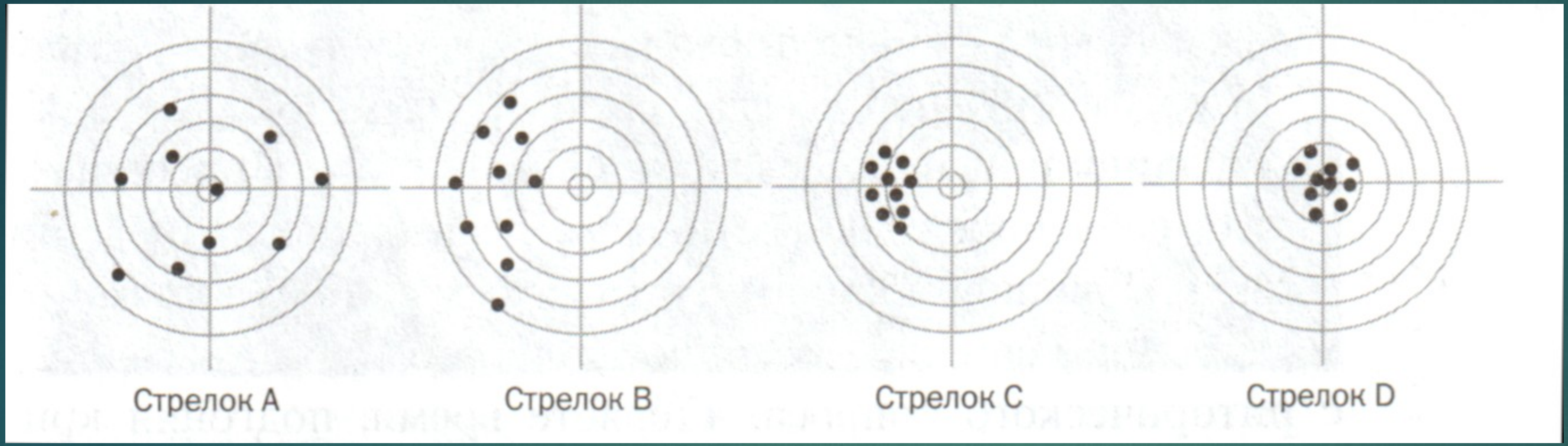
- Стационарность $P_k(t) = f(k,t)$
- Ординарность (1 в Δt)
- Отсутствие последействия

λ = 3 мин$^{-1}$ t = 2 мин k = 4 шт.

$$P_t(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

$$P_2(4) = \frac{(6)^4}{4!} e^{-6} = \frac{36 * 36}{24 * 408,18} = 0,135$$

Планирование и проведение эксперимента



Состоятельность оценки
Эффективность оценки
Достаточность оценки

Вывод о целом на основании частного

N взаимно независимых случайных величин $X_1, X_2, \cdots, X_n$

$$M(\bar{X}) = \frac{X_1 + X_2 + .. + X_n}{n} = a$$

$$D(\bar{X}) = D(\frac{X_1 + X_2 + .. + X_n}{n}) = \frac{D(X_1) + D(X_2) + .. + D(X_n)}{n} = \frac{nD}{n^2} = \frac{D}{n}$$

$$\sigma = \sqrt{D}$$

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

Неравенство Чебышёва

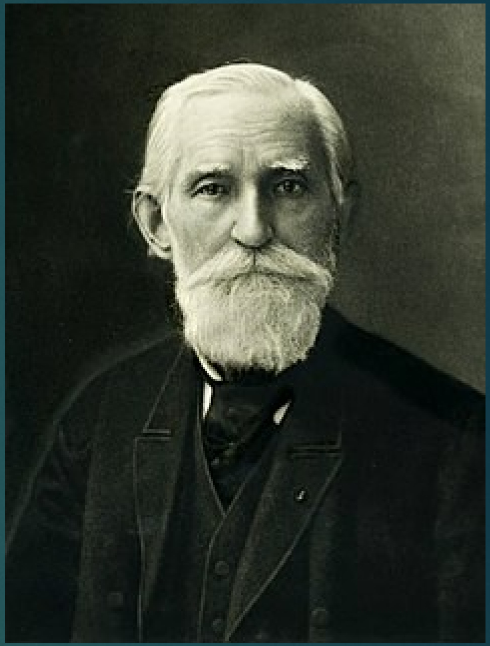
$P(| X — M(X)| < \varepsilon) \geq 1 - D(X)/\varepsilon^2$ а

$P(| X — M(X)| < \varepsilon) + P(| X — M(X)|) \geq \varepsilon = 1$

$D(X) = |x_1 - M(X)|^2 p_1 + |x_2 — M(X)|^2 p_2 . . . + |x_n — M(X)|^2 p_n$

$D(X) > |x_k — M(X)|^2 p_k . . . + |x_n — M(X)|^2 p_n =$

$= \varepsilon^2 (p_k + p_{k+1} + . . . + p_n)$



Пафнутий Львович Чебышёв
1821 -1894

$P(| X — M(X)|) \geq \varepsilon$

==> $D(X) \geq \varepsilon^2 P(| X — M(X)|) \geq \varepsilon$

$P(| X — M(X)| \geq \varepsilon) \leq D(X) / \varepsilon^2$ ==> $P(| X — M(X)| < \varepsilon) \geq 1 - D(X) / \varepsilon^2$

Теорема Чебышёва

Дисперсии ограничены С

а

$$\left|\frac{X_1+X_2+..+X_n}{n}-\frac{M(X_1)+M(X_2)+..+M(X_n)}{n}\right|\leq\epsilon$$

$$P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}\right|<\varepsilon\right)\geqslant 1-\frac{C}{n\varepsilon^2}.$$

Теорема Бернулли

$$\lim_{n\to\infty}P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=1$$

X_1, X_2, X_n - попарно независимые случайные величины , имеющие одно и то же математическое ожидание а, и если дисперсии этих величин равномерно ограничены , то, как бы мало не было число ε, вероятность неравенства будет как угодно близка к единице, если число случайных величин достаточно велико .

Кейс: случайна ли судебная ошибка?

~~Заказной отчет. Оценщик.~~

Преследуемые Должники

Кредиторы, «оставленные с носом»

Как будто торги с 1 участником и победителем

Финуправляющий

Регистрация сообщ. о преступлении в Сургуте и Н-ске в КУСП

Присвоение 6 тыс. Госсредств и 120 тыс. представителя

Как будто Цессия

1080% прибыльность!

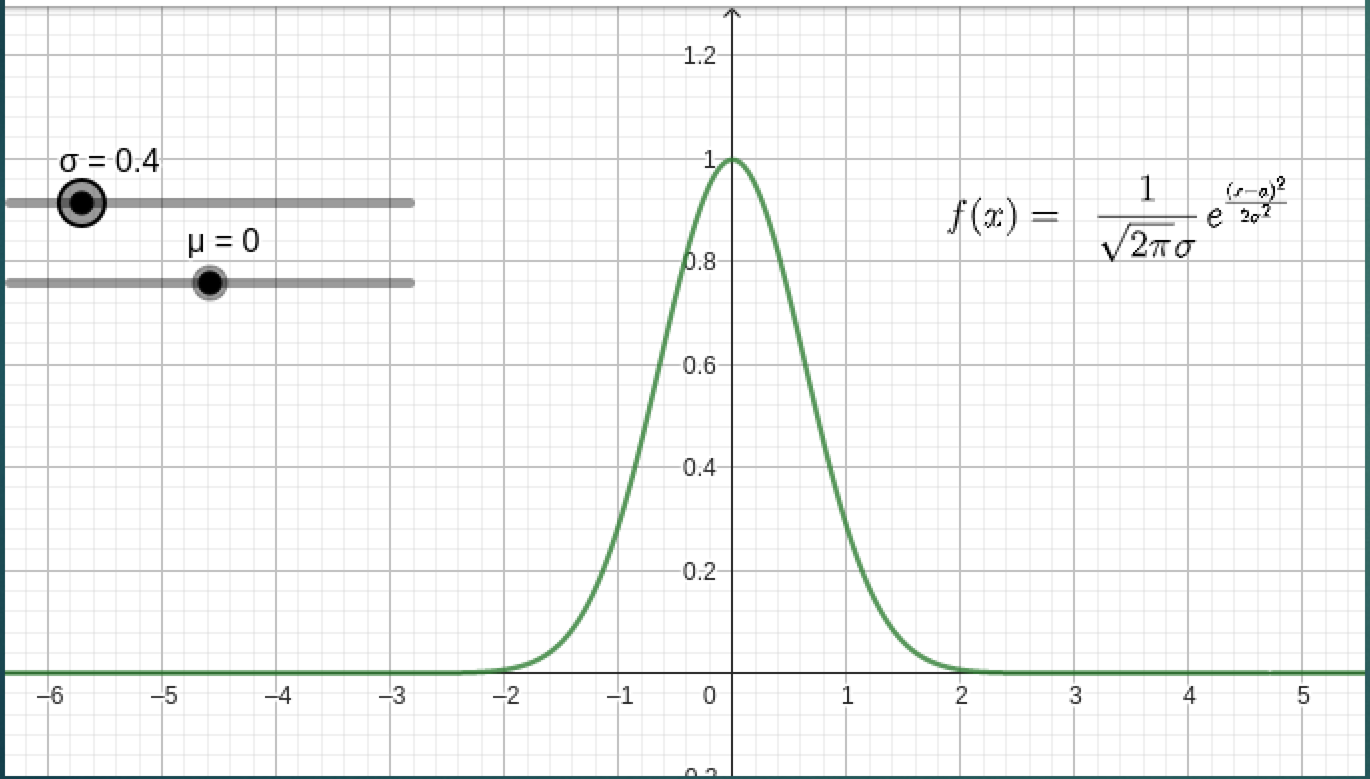
Как будто Продажа

«Соратница» Финуправляющего

Статистические гипотезы

Нулевая гипотеза H_0 о том, *что нарушения Законности в деле No арбитражного суда имеют* случайный *характер*.

Альтернативная гипотеза H1: отклонения Законности судебных актов по делу No арбитражного суда имеют *систематический* характер.



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

Эталонное решение не содержит ошибок μ = 0.
Стандартное отклонение σ = 0.78

«80% судебных актов не содержит ни одной ошибки».

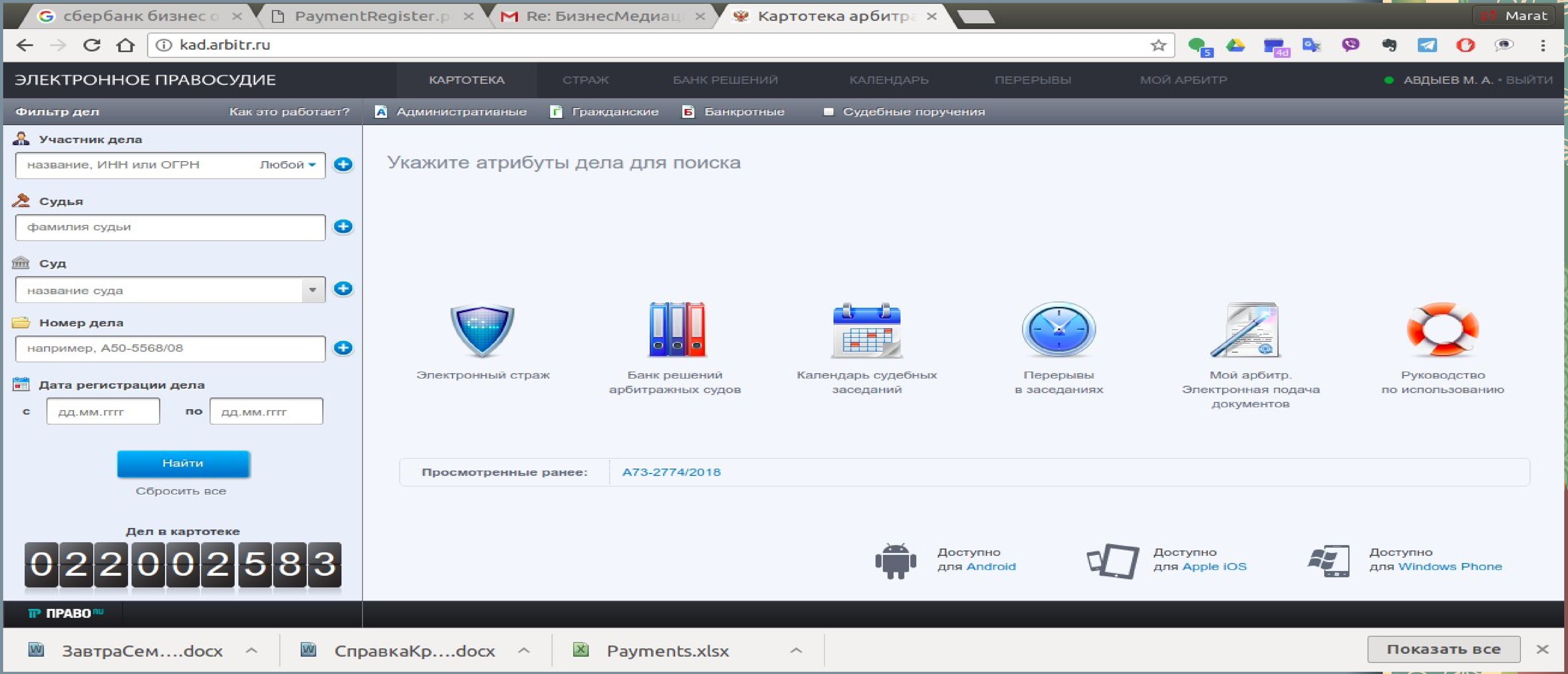
Анализ

Производственный контроль качества продукта

Постановление Пленума Верховного Суда Российской Федерации от 19 декабря 2003 г. N 23 г. «О судебном решении», ***АПК РФ, ФЗ***

Как есть	Как должно быть
Подложный документ	Исключается
Заведомо ложное экспертное заключение	Стандарты оценки
Игнорирование общеизвестных доказательств	Не допустимо

Генеральная совокупность



Оценка вероятности H_1

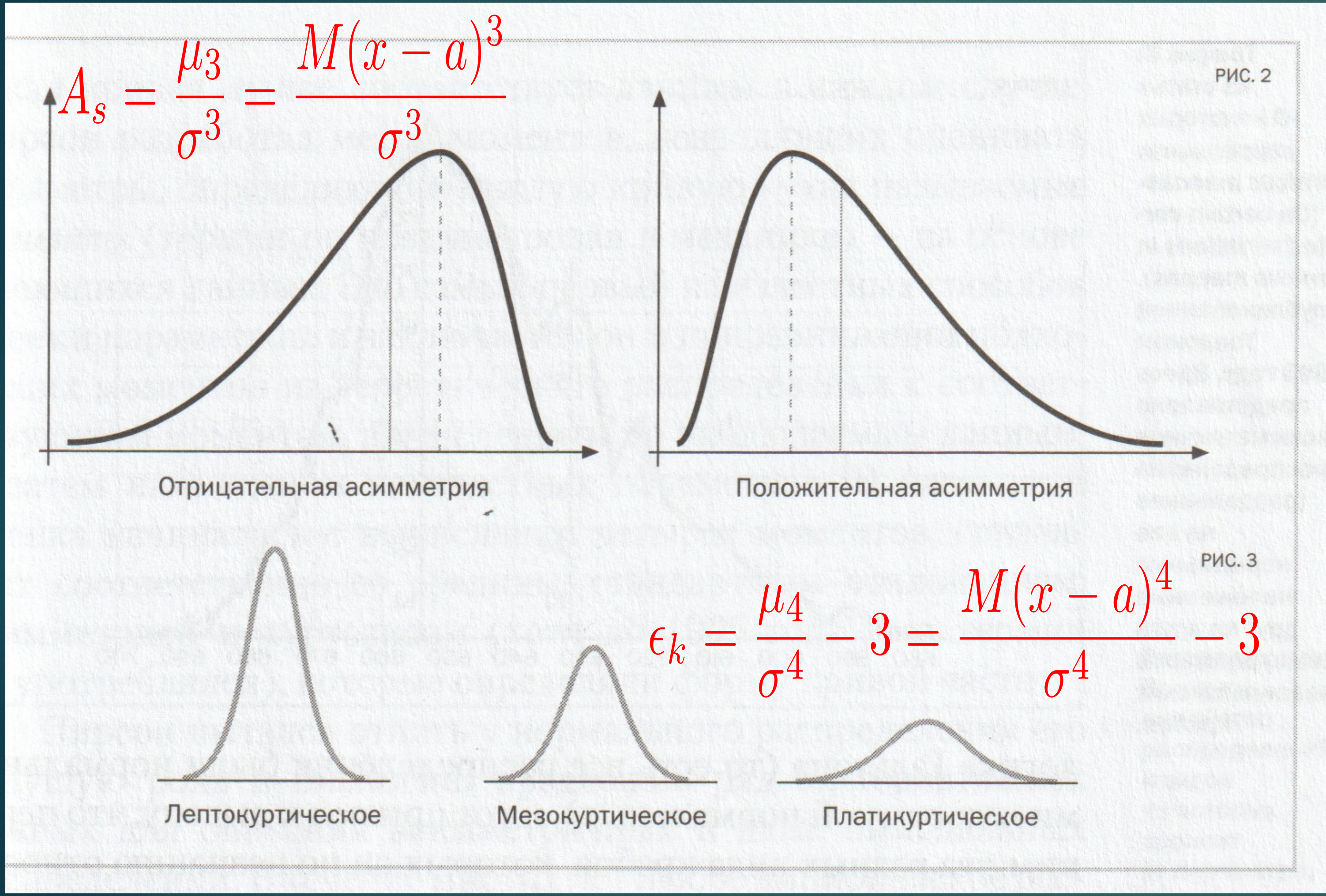
0	1
1	1 1
2	1 2 1
3	1 3 3 1
4	1 4 6 4 1
5	1 5 10 10 5 1
6	1 6 15 20 15 6 1
7	1 7 21 35 35 21 7 1
8	1 8 28 56 70 56 28 8 1

$$P_n(m) = C_n^m * p^m * q^{n-m}$$

$$P = \left(\frac{1}{5}\right)^{12} = \frac{1}{2^{12\ln(5)/\ln(2)}} = \frac{1}{2^{27.8}} = \frac{1}{1024 * 1024 * 256} \approx$$

$4 * 10^{-9}$

Асимметрия. Куртозис. Моменты



$$A_s = \frac{\mu_3}{\sigma^3} = \frac{M(x-a)^3}{\sigma^3}$$

$$\epsilon_k = \frac{\mu_4}{\sigma^4} - 3 = \frac{M(x-a)^4}{\sigma^4} - 3$$

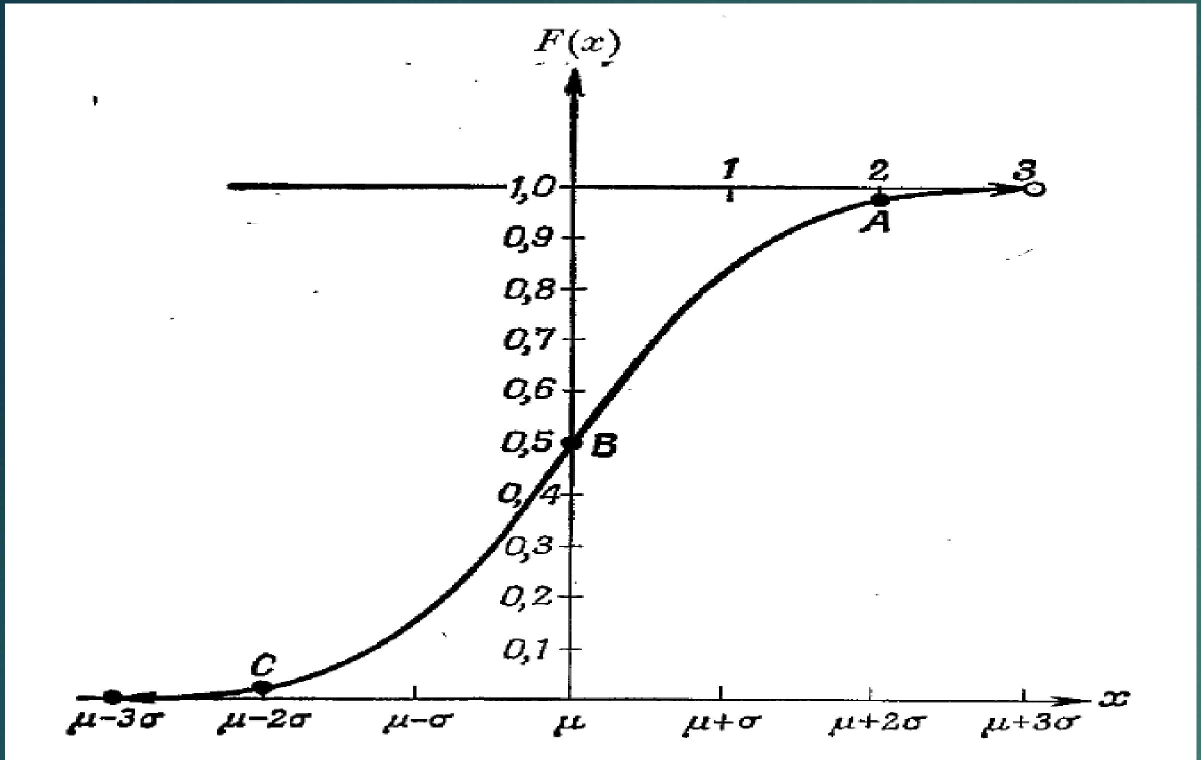
$$\nu_k = M(x^k)$$

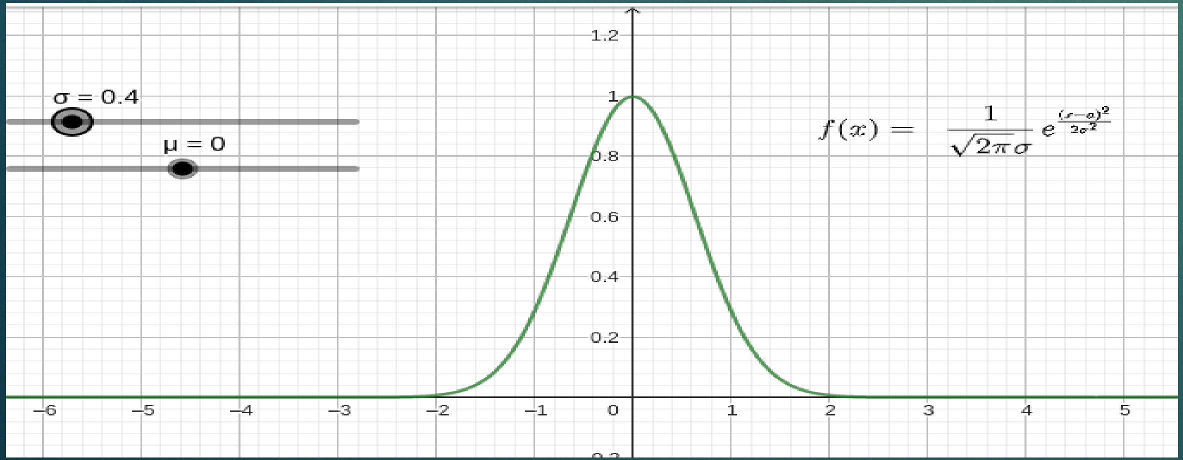
$$\nu_2 - \nu_1^2 = D$$

$$\mu_1 = M\left\{(x - M(x))\right\} = 0$$

$$\mu_2 = M\left\{(x - M(x))^2\right\} = D$$

Функция нормального распределения F





$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{s^2}{2}} ds$$

Плотность вероятности

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

F(x) = Φ(x) +0,5

Фишер. Критерий значимости.

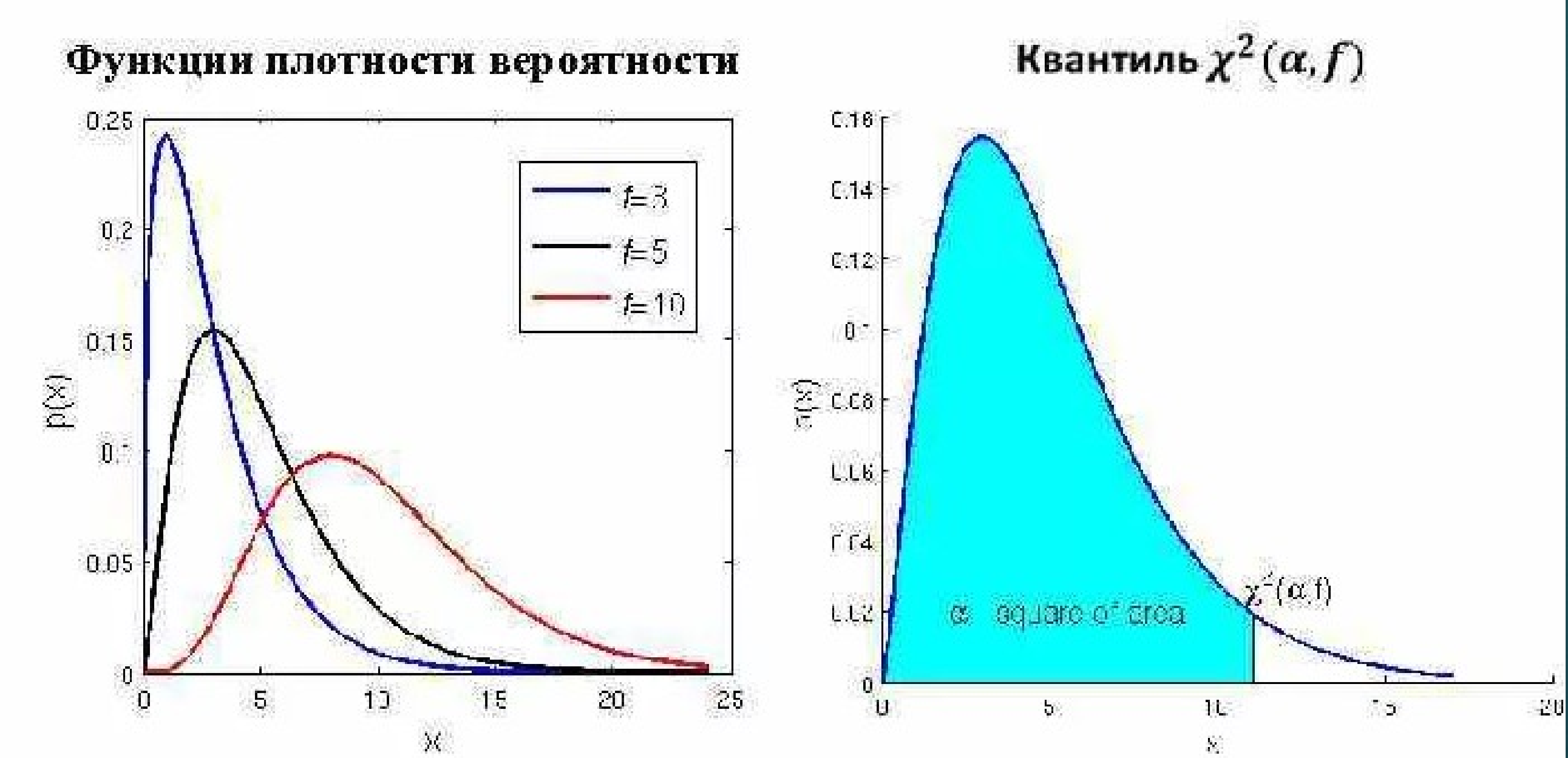
Если наблюдаемая гипотеза H_0, то наблюдаемые данные статистически
незначимы для $\alpha = 0,05$ с высокой вероятностью $1- \alpha = 0,95$.

Наблюдаемая выборка X статистически значима на уровне $\alpha = 0,05$, следовательно гипотеза H_0 неверна.

Отвергая справедливую H_0	Ошибка первого рода
Принимая вместо справедливой конкурирующей H_1	Ошибка второго рода

χ^2 с k степенями свободы

$$\chi^2 = \sum_{1}^{k} X_i^2 \qquad X \in N(0,1)$$



Как рассчитывается χ^2

$$\chi^2 = \sum_1^k X_i^2$$

$$\chi_*^2 = \sum_1^k (n_i - \tilde{n}_i)^2 / \tilde{n}_i = \sum_1^k n_i^2 / \tilde{n}_i - n$$

$k = s - 1 - r$

где s - число групп (частичных интервалов) выборки
r - число параметров предполагаемого распределения, которые оценены по данным выборки

Сравнение	Нулевая гипотеза
$\chi^2 < \chi^2_{кр}$	Принимается
$\chi^2 > \chi^2_{кр}$	Отвергается

$\chi^2 < \chi^2_{кр}(\alpha, k)$

Проверка нулевой гипотезы

Число погибших офицеров x	0	1	2	3	4	5	6	Всего
Число донесений r_x указанием x погиб.	109	65	22	3	1	0	0	200
$x * r_x$	0	65	44	9	4	0	0	122
Ожидаемые частоты (по Пуассону) част. m_x	108,7	66,3	20,2	4,1	0,6	0,07	0,01	199,98

$$\sum r_x = 200 \ldots \sum x \cdot r_x = 122$$

x	0	1	2	>= 3
rx	109	65	22	4
x * rx	0	65	44	13
mx	108,7	66,3	20,2	4,77
$(r_x — m_x)^2/m_x$	0,001	0,025	0,160	0,124

0,311

$$m_j = n * p_j = 200\frac{0,61^j}{j!}e^{-0,61}$$

k = 4 - 1 -1 = 2

$X^2_{кр}$ (0,05, 2) = 6,0

Доверительный интервал, σ известно

$$\bar{X} = a \qquad \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

$$P(|\ \bar{X} - a|) < t\sigma/\sqrt{n} = 2\Phi(t)$$

$$P(|\ \bar{X} - a|) < \delta = \gamma$$

$$\mathrm{t} = \delta\frac{\sqrt{n}}{\sigma}$$

$$P(|\ \bar{X} - a|) < \delta = 2\Phi\left(\frac{\delta}{\sigma}\right) = 2\Phi(t)$$

$$P(\bar{x} - t\sigma/\sqrt{n} < a < \bar{x} + t\sigma/\sqrt{n} = 2\Phi(t) = \gamma$$

С надежностью γ можно утверждать, что ЭТОТ доверительный интервал покрывает неизвестный параметр *а* с точность оценки δ = tσ√n

Доверительный интервал N(a = ?, σ)

$$\bar{X} = a \qquad \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

$$P(|\bar{X} - a|) < t\sigma/\sqrt{n} = 2\Phi(t)$$

$$P(|\bar{X} - a|) < \delta = \gamma$$

$$\mathrm{t} = \delta\frac{\sqrt{n}}{\sigma}$$

$$P(|\bar{X} - a|) < \delta = 2\Phi\left(\frac{\delta}{\sigma}\right) = 2\Phi(t)$$

$$P(\bar{x} - t\sigma/\sqrt{n} < a < \bar{x} + t\sigma/\sqrt{n} = 2\Phi(t) = \gamma$$

С надежностью γ можно утверждать, что ЭТОТ доверительный интервал покрывает неизвестный параметр *a* с точность оценки δ = tσ√n

Доверительный интервал N(a = ?, σ?)

$$P(|\bar{X} - a|) < \delta = \gamma$$

$$P(|\bar{X} - a|) < t\sigma/\sqrt{n} = 2\Phi(t)$$

$$T = \frac{\bar{X} - a}{\frac{S}{\sqrt{n}}}$$

Случайная величина

$$P\left(\left|\frac{\bar{X} - a}{\frac{S}{\sqrt{n}}} < \gamma\right|\right) = 2\int_o^{t_\gamma} S(t,n)dt = \gamma$$

$$\bar{x} - t_\gamma S/\sqrt{n} < a < \bar{x} + t_\gamma S/\sqrt{n}$$

Пользуясь распределением Стьюдента, мы нашли ЭТОТ доверительный интервал, покрывающий неизвестный параметр a с надёжностью γ

Корреляция и регрессия

$$\overline{xy} = \frac{\sum_1^k xyn_{xy}}{n} \qquad \bar{y} = \frac{\sum_1^k yn_y}{n} \qquad \bar{x} = \frac{\sum_1^k xn_x}{n}$$

$$\sigma_x^2 = D(x) = \overline{x^2} - \bar{x}^2 \qquad \sigma_y^2 = D(y) = \overline{y^2} - \bar{y}^2$$

Y	X					n_y
	1	2	3	4	5	
28	0	0	0	37	3	**40**
38	0	0	13	6	0	**19**
48	0	13	10	0	0	**23**
58	17	1	0	0	0	**18**
n_x	**17**	**14**	**23**	**43**	**3**	**100**

Корреляция и регрессия

D(x) = 10,43-9,06 =1,37

$$\bar{x} = \frac{1*17+2*14+3*23+4*43+5*3}{100} = 3,01 \Longrightarrow \bar{x}^2 = 9,06$$

σ_x = 1,17

$$\bar{x^2} = \frac{1^2*17+2^2*14+3^2*23+4^2*43+5^2*3}{100} = 10,43$$

<y>= **39,90**
<y>^{2}= **1592**
<y^2>= 1723,4

D(y) = 131,39
σ_y = 11,46
<xy>= **107,48**

Y	X					n_y
	1	2	3	4	5	
28	0	0	0	37	3	**40**
38	0	0	13	6	0	**19**
48	0	13	10	0	0	**23**
58	17	1	0	0	0	**18**
n_x	**17**	**14**	**23**			

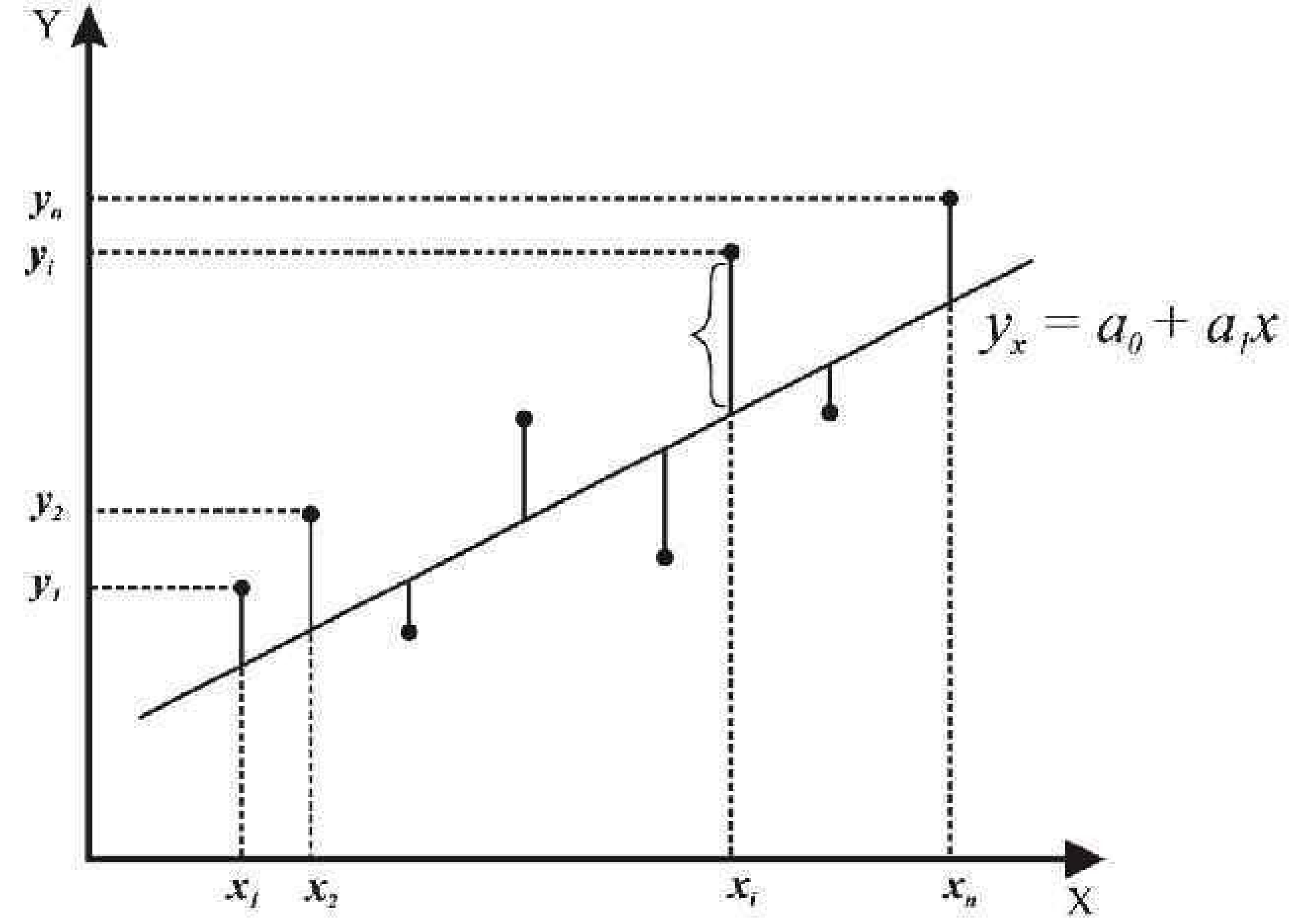
$$r = \frac{\overleftrightarrow{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y} = \frac{107,48 - 3,01 \cdot 39,9}{1,17 \cdot 11,46} = -0,94$$

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x} \cdot (x - \bar{x})$$

$$\bar{xy} = \frac{17 \cdot 1 \cdot 58 + 13 \cdot 2 \cdot 48 + 1 \cdot 2 \cdot 58 + 13 \cdot 3 \cdot 38 + 10 \cdot 3 \cdot 48 + 37 \cdot 4 \cdot 28 + 6 \cdot 4 \cdot 38 + 3 \cdot 5 \cdot 28}{100} = 107,48$$

Линейная регрессия

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x} \cdot (x - \bar{x})$$



Линейная регрессия

Расчет коэфф. корреляции мальчиков 2-15 лет.

$$\mu_{xy} = M\{[X - M(X)][Y - M(Y)]\}$$

$$\mu_{xy} \leq \sqrt{D_x D_y)} \qquad |\mu_{xy}| \leq \sqrt{D_x D_y)} = \sigma_x \sigma_y$$

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y} \qquad |r_{xy}| \leq 1$$

Лет	Вес (кг)	Рост (см.)
2	13	86
3	13	93
4	15	102
5	18	102
6	20	114
7	21	120
8	23	124
9	25	129
10	27	135
11	31	140
12	35	146
13	40	152
14	44	158
15	53	163

Calculator Suite - GeoGebra | ЕКП | Data.gov.ru

data.gov.ru | ЕКП | Data.gov.ru

Data.gov.ru
открытые данные России

Данные Библиотека Сообщества Сервисы

Вход

→ Данные → Наборы данных

Набор данных | Журнал

ЕКП

Спорт 15032 1488

-25

Организация-публикатор:
Министерство спорта Российской Федерации

Посмотреть на карте

Добавлено:
Тимко Валентина Ивановна, консультант отдела физической культуры и массового спорта

Телефон: +74992676906

Адрес электронной почты:
timko@minsport.gov.ru

Дата первой публикации набора данных:
10.06.2013

Дата последнего внесения изменений: 14.09.2015

Условия использования открытых данных

Актуальная версия

Просмотреть данные

Дата	Автор	Источник	Формат	Действия
14.09.2015	Служба технической поддержки	Внешний	csv(16.36 МБ)	Скачать

Последний набор (Windows)
Последняя структура (Windows)
Последний набор (UTF-8)
Последняя структура (UTF-8)

Паспорт набора

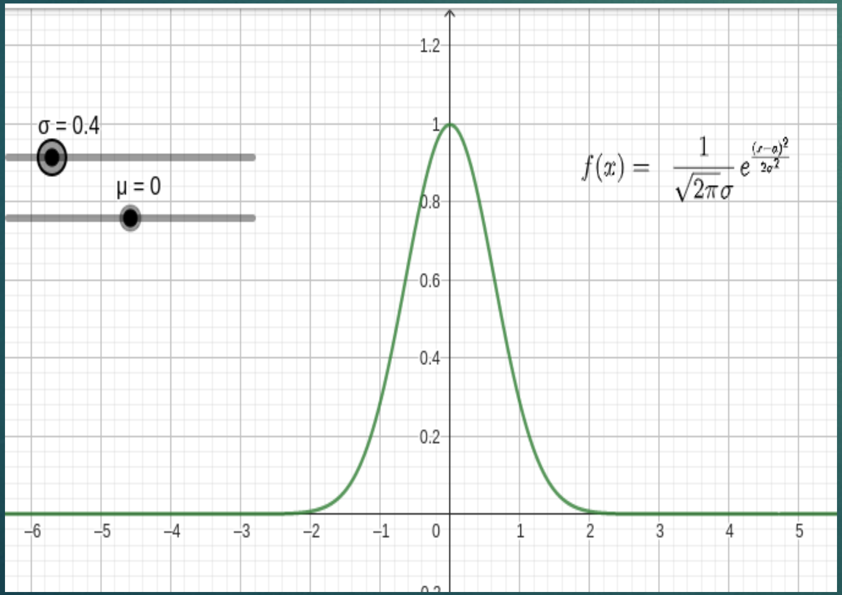
Дополнительные файлы (1)

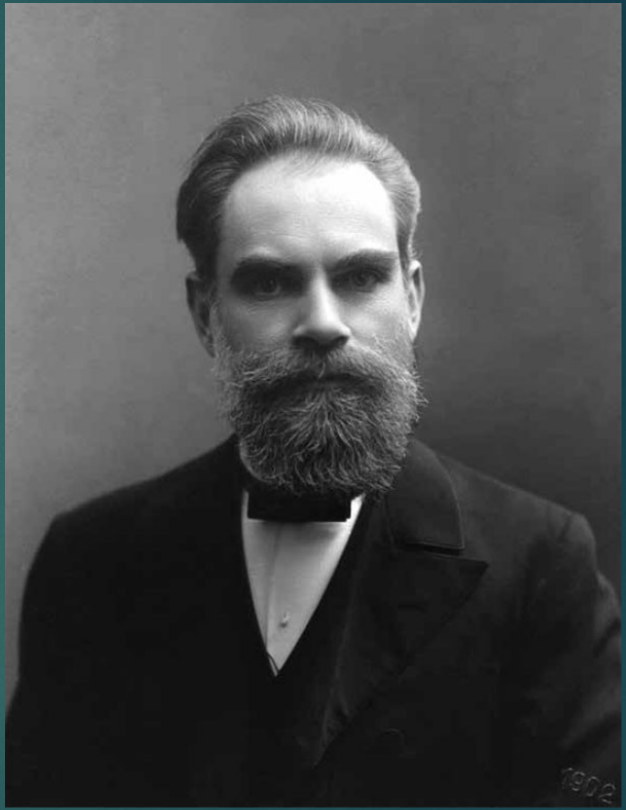
Дата	Название документа	Формат	Закачать
14.09.2015	Оригинал файла набора открытых данных	csv(15.62 МБ)	Скачать

Для того, чтобы оставить комментарий, Вам нужно зарегистрироваться

Нашли ошибку? Расскажите нам!

Закон больших чисел. Центральная предельная теорема Ляпунова.





ANOVA дисперсионный анализ

$$mean = \frac{11+6+1+9+5+3}{6} = 5,83$$

Analysis of Variance

Средство A	Средство B	Средство C
11	6	1
9	5	3
<10>	<5,5>	<2>

Число степеней свободы
Всего налблюдений во всех трёх группах 6
6 — 1 = 5 степеней свободы

Всего групп 3

3 — 2 = 1 степеней свободы

Разность: кол-во групп — кол-во наблюдений

5 — 3 = 2 степеней свободы

«Эта разница статистически значима?

ANOVA дисперсионный анализ Пусть $\alpha = 0{,}05$

$SSA = 2\{ (10-5{,}83)^2 + (5{,}5-5{,}83)^2 + (2-5{,}83)^2 \} = 64{,}33$

$SSW = (11-10)^2 + (9--10)^2 + (6--5{,}5)^2 + (5--5{,}5)^2 + (1--2)^2 + (3--2)^2 = 4{,}5$

$SST = (11-5{,}83)^2 + (6-5{,}83)^2 + (1-5{,}83)^2 + (9-5{,}832)^2 + (5-5{,}83)^2 + (3-5{,}83)^2 = 68{,}83$

$SST = 68{,}83 = 64{,}33+4{,}5 = SSA + SSW$

A	B	C
11	6	1
9	5	3
<10>	<5,5>	<2>

Таблица ANOVA

Источник вариации	Сумма квадратов	Степеней свободы	Среднее квадратов	$F(k_1, k_2, \alpha)$
SSA Между группами	64,33	2	32,17	21,44
Внутри групп SSW	4,5	3		
Общая SST вариация	68,83	5		

Отношение большего к меньшему 19,16