

Вычисление числовых характеристик случайных величин

Построение статистической таблицы

(*гистограмма*;

полигон накопленных частот – *кумулята*

и т. д.)

1-й этап

Составляем таблицу первичных данных

располагаем данные
в порядке их поступления

Например,

Результаты тестирования

50 испытуемых

16,2 15,4 15,3 15,3 15,3 15,4 16,8 17,8 16,2 15,9
15,5 14,5 16,0 15,5 15,8 14,7 16,0 15,6 15,5 15,0
14,3 14,8 13,7 14,8 14,2 12,8 14,6 15,0 13,6 14,2
16,6 16,1 16,1 14,2 15,8 16,9 15,6 15,6 16,4 16,4
15,8 15,8 16,2 16,2 14,2 15,0 16,1 15,0 15,2 14,2

$$\bar{x}_{50} = \frac{1}{50} (12,8 + 15,4 + 15,3 + \dots + 14,2) = \frac{771,2}{50} \approx 15,4$$

2-й этап

Находим *минимальное* и *максимальное*
значение (из таблицы первичных данных)

50 испытуемых

16,2 15,4 15,3 15,3 15,3 15,4 16,8 17,8 16,2 15,9
15,5 14,5 16,0 15,5 15,8 14,7 16,0 15,6 15,5 15,0
14,3 14,8 13,7 14,8 14,2 12,8 14,6 15,0 13,6 14,2
16,6 16,1 16,1 14,2 15,8 16,9 15,6 15,6 16,4 16,4
15,8 15,8 16,2 16,2 14,2 15,0 16,1 15,0 15,2 14,2

3-й этап

Находим *число интервалов*

(из объема выборки)

**50 испытуемых = объем
выборки**

Формула
Стерджеса:

$$k = 1 + 3,32 \cdot \lg n$$

Объем выборки (n)	Число интервалов (k)
25 - 40	5 - 6
40 - 60	6 - 8
60 - 100	7 - 10
100 - 200	8 - 12
Больше 200	10 - 15

4-й этап

Находим *ширину интервала (h)*

По формуле:

$$h = \frac{X_{max} - X_{min}}{k}$$

h — ширина интервала

k — число интервалов

Для нашего случая
ширина интервала (h)

равна:

$$h = \frac{17,8 - 12,8}{7} = 0,714 \approx 0,8$$

5-й этап

Находим *нижние границы интервалов* (X_n)

$$x_{н1} = x_{min} - \frac{h}{2} = 12,8 - \frac{0,8}{2} = 12,4$$

$$x_{н2} = x_{н1} + h = 12,4 + 0,8 = 13,2$$

$$x_{н3} = x_{н2} + h = 13,2 + 0,8 = 14,0$$

$$x_{н4} = x_{н3} + h = 14,0 + 0,8 = 14,8$$

$$x_{н5} = x_{н4} + h = 14,8 + 0,8 = 15,6$$

$$x_{н6} = x_{н5} + h = 15,6 + 0,8 = 16,4$$

$$x_{н7} = x_{н6} + h = 16,4 + 0,8 = 17,2$$

6-й этап

Находим **верхние границы интервалов (Хв)**

$$x_{в1} = x_{н2} - ПВ = 13,2 - 0,1 = 13,1$$

где ПВ = погрешность вычисления = 0,1

$$x_{в2} = x_{н3} - ПВ = 14,0 + 0,1 = 13,9$$

$$x_{в3} = x_{н4} - ПВ = 14,8 - 0,1 = 14,7$$

$$x_{в4} = x_{н5} - ПВ = 15,6 + 0,1 = 15,5$$

$$x_{в5} = x_{н6} - ПВ = 16,4 - 0,1 = 16,3$$

$$x_{в6} = x_{н7} - ПВ = 17,2 - 0,1 = 17,1$$

$$x_{в7} = x_{н8} - ПВ = 18,0 - 0,1 = 17,9$$

7-й этап

Находим *срединные значения интервалов* (X_{cp})

$$x_{cp1} = \frac{x_{в1} + x_{н1}}{2} = 12,1$$

$$x_{cp2} = 13,6$$

$$x_{cp3} = 14,4$$

$$x_{cp4} = 15,2$$

$$x_{cp5} = 16,0$$

$$x_{cp6} = 16,8$$

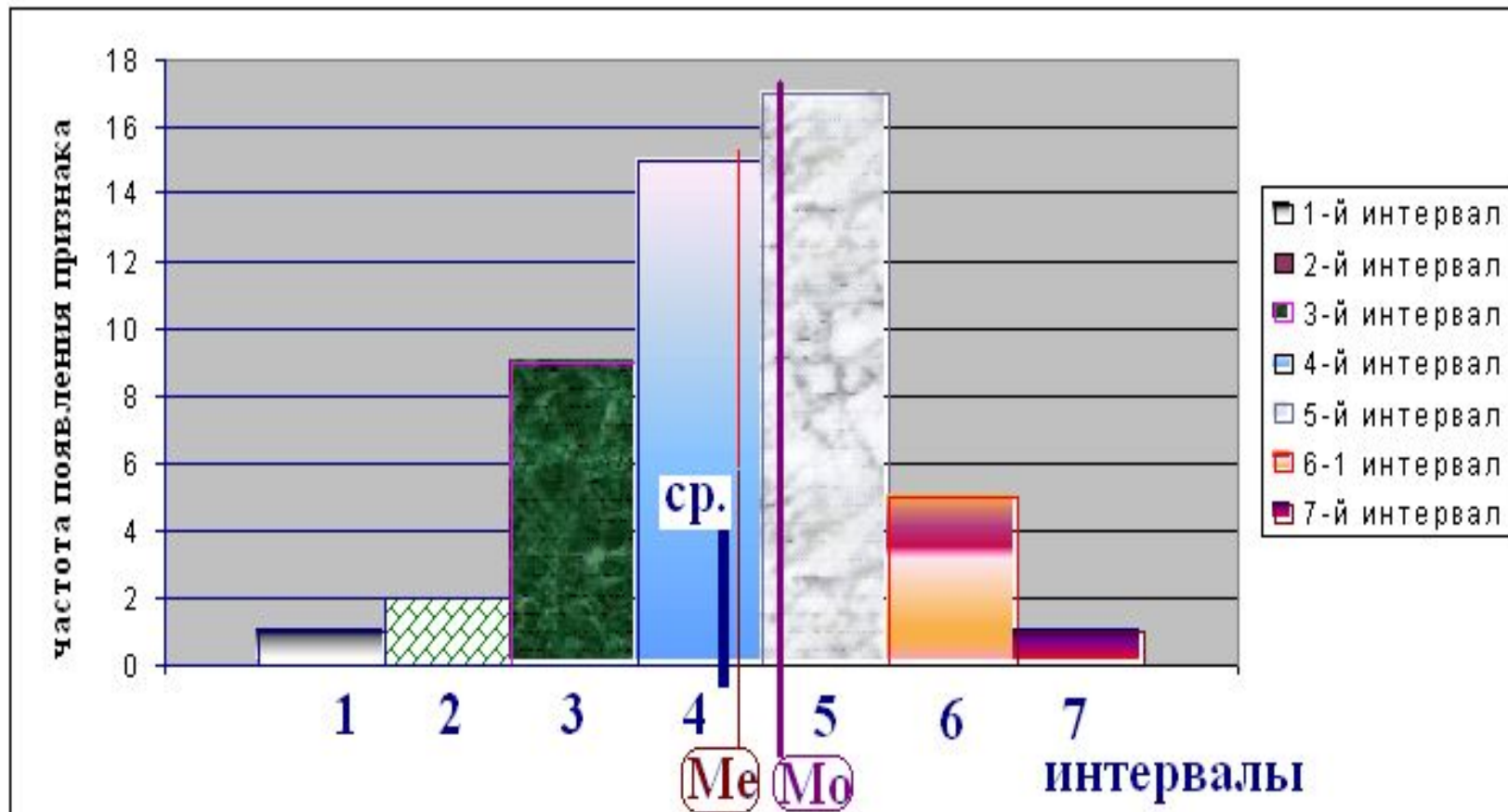
$$x_{cp7} = 17,6$$

8-й этап

Находим частоту интервала (n_i)

№ интервала	Границы интервалов	Срединные значения интервалов	Распределение данных на интервалах	Частоты (n_i)
1	12,4-13,1	12,8	/	1
2	13,2-13,9	13,6	//	2
3	14,0-14,7	14,4	////////	9
4	14,8-15,5	15,2	//////////	15
5	15,6-16,3	16,0	////////////////	17
6	16,4-17,1	16,8	////	5
7	17,2-17,9	17,6	/	1
сумма				50

Гистограмма



Полигон частот

(эмпирическое распределение данных)



Полученной кривой можно поставить
в соответствие формулу: $Y = f(x)$

Для этого используют программные пакеты:

Curve-expert

Math Lab

Origin

Maple

Advanced Graphic

Model Information - [Gaussian Model]

Coeficients History Covariance Residuals Comments

Gaussian Model

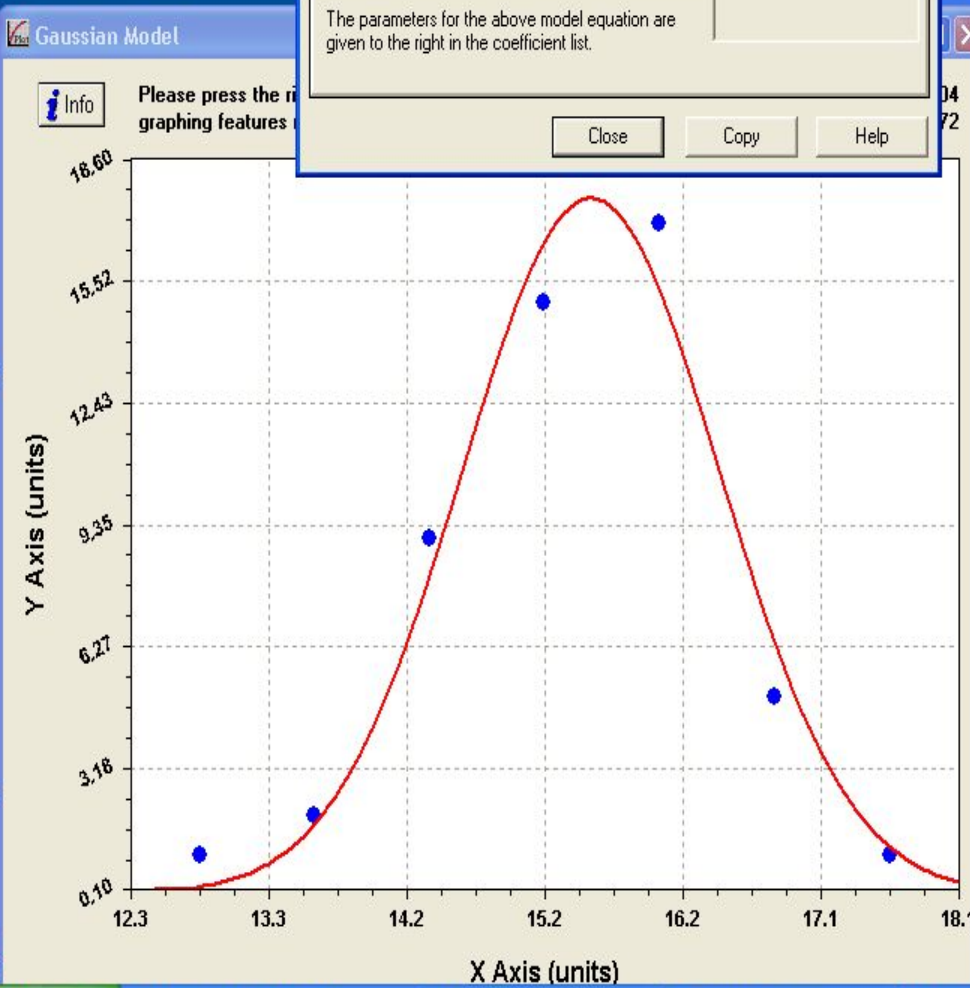
$$y = ae^{-\frac{(x-b)^2}{2c^2}}$$

The parameters for the above model equation are given to the right in the coefficient list.

Coeficients:

a = 17.621139
 b = 15.519069
 c = 0.89716084

Close Copy Help



CurveExpert 1.3 - [UNTITLED.DAT*]

File Edit Data Interpolate Apply Fit Tools Window Help

	X	Y
1	12.8	1
2	13.6	2
3	14.4	9
4	15.2	15
5	16	17
6	16.8	5
7	17.6	1
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		

Rank Regression

- Gaussian Model
- Sinusoidal Fit
- Vapor Pressure
- Polynomial Fit
- Quadratic Fit
- Gompertz Relation
- Hyperbolic Fit
- Logarithm Fit
- Modified Exponential
- Modified Geometric
- Linear Fit
- User-Defined Model
- Power Fit
- Binomial Logarithm

Data Plot

9-й этап

Находим *среднее по генеральной совокупности – математическое ожидание*

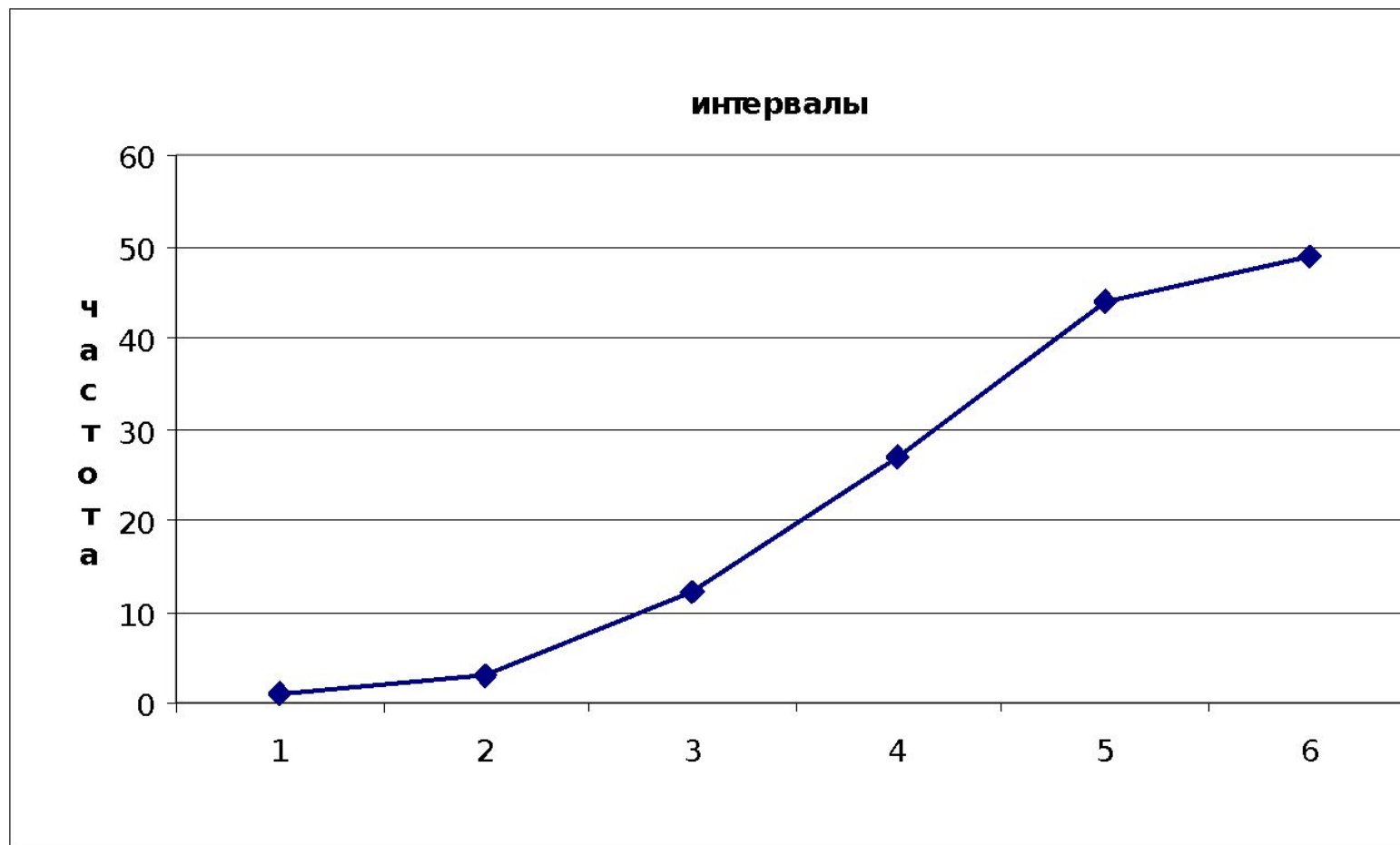
$$x_c = \bar{x} = \int x f(x) dx$$

10-й этап

Находим накопленную частоту интервала (n_{xi})

№ интервала	Границы интервалов	Срединные значения интервалов	Распределение данных на интервалах	Частоты (n_i)	Накопленные частоты (n_{xi})
1	12,4-13,1	12,8	/	1	1
2	13,2-13,9	13,6	//	2	3
3	14,0-14,7	14,4	////////	9	12
4	14,8-15,5	15,2	//////////	15	27
5	15,6-16,3	16,0	////////////////	17	44
6	16,4-17,1	16,8	////	5	49
7	17,2-17,9	17,6	/	1	50
сумма				50	

КУМУЛЯТА (полигон накопленных частот)



Пример 2

Результаты тестирования

50
испытуемых

<i>№n/n</i>	<i>x_i</i>	<i>n_i</i>	<i>x_in_i</i>
1	12,8	1	12,8
2	13,6	2	27,2
3	14,4	9	129,6
4	15,2	15	228,0
5	16,0	17	272,0
6	16,8	5	84,0
7	17,6	1	17,6
<i>сумма</i>			771,2

$$\bar{x}_{50} = \frac{1}{50} (12,8 + 27,2 + 129,6 + 228,0 + \dots + 17,6) = \frac{771,2}{50} \approx 15,4$$

Conclusion and OUTLOOK

Т.Е. Итак,

1. Вы, конечно же, поняли,

Создать статистическую таблицу

А затем, исходя из нее, получить эмпирическое распределение
(гистограмма, полигон частот, кумулята)

Спасибо за внимание