

Тема 6. Математическая статистика

Основные положения:

Пусть из генеральной совокупности извлечена выборка, причем

x_1 наблюдалось n_1 раз,

x_2 - n_2 раз,

x_k - n_k раз,

$\sum n_i = n$ - объем выборки.

Числа наблюдений называют частотами, а их отношение к объему выборки

$\frac{n_i}{n} = w_i$ - относительными частотами. Причем $\sum_{i=1}^k w_i = 1$, т.е. сумма

относительных частот всех вариантов равна единице.

Полигоном частот называют ломаную, отрезки которой соединяют точки

$(x_1, n_1), (x_2, n_2), \dots, (x_k, n_k)$. Для построения полигона частот на оси абсцисс

откладывают варианты x_i , а на оси ординат – соответствующие им частоты n_i . Точки (x_i, n_i) соединяют отрезками прямых и получают полигон частот.

Полигоном относительных частот называют ломаную, отрезки которой соединяют точки $(x_1, w_1), (x_2, w_2), \dots, (x_k, w_k)$.

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы

длиною h , а высоты равны отношению $\frac{n_i}{h}$.

Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные

интервалы длиною h , а высоты равны отношению $\frac{w_i}{h}$.

Назначения подпакетов, вызываемых with(stats):

anova — дисперсионный анализ;

describe — вычисления числовых характеристик выборок;

fit — регрессионный анализ;

random — генерация случайных чисел с заданным законом распределения;

statevalf — теория вероятностей;

statplots — графические представления выборок;

transform — группировка (преобразование) выборочных данных.

Необходимость статистической обработки данных возникает тогда, когда есть некоторый массив данных-выборка. Если нужно получить случайную выборку, то используют пакет

random[distribution](quantity, uniform, method).

Параметры подпакета:

distribution — закон распределения выборки,

quantity — объем выборки,

uniform — использовать преобразование случайного равномерного распределения в данный закон,

default — по умолчанию.

Устанавливаем, чтобы не рябило в глазах, число значащих цифр равное трем:

```
> Digits:=3 ;
```

Пример:

```
> restart;
```

```
> with(stats):
```

```
> Digits:=3;
```

Генерируем выборку объема $n = 50$, имеющую нормальный закон распределения с математическим ожиданием $\mu = 10$ и среднеквадратическим отклонением $\sigma = 2$:

```
> x:=[random[normald[10,2]](50,'default','inverse')];  
δ:= [9.63, 9.07, 9.19, 9.87, 10.3, 11.3, 6.30, 11.2, 10.5, 11.3, 8.71, 9.01, 11.7, 6.48, 7.30,  
13.5, 11.8, 9.77, 10.7, 12.8, 13.3, 7.90, 7.97, 9.64, 10.1, 8.79, 8.45, 10.9, 12.0, 10.9,  
8.84, 11.6, 11.4, 10.7, 9.03, 4.96, 7.12, 10.7, 7.77, 10.9, 7.80, 14.8, 9.76, 11.3, 10.0,  
8.31, 11.9, 14.8, 7.62, 9.18]
```

Типовой пример. В результате эксперимента получены данные, записанные в виде статистического ряда.

9.19, 11.5, 10.7, 12.6, 13.0, 12.3, 7.46, 8.92, 8.80, 11.6, 11.9, 10.9, 5.82, 8.89,
9.32, 8.30, 8.76, 8.01, 15.5, 12.3, 9.46, 9.11, 12.1, 12.5, 9.33, 11.0, 10.1, 9.61,
13.7, 15.0, 12.2, 13.1, 11.7, 10.4, 11.5, 9.02, 9.23, 7.16, 12.0, 10.6, 6.39, 6.97,
9.03, 6.84, 8.29, 10.5, 11.7, 7.05, 12.1, 9.53.

Требуется:

- записать значения результатов эксперимента в виде вариационного ряда;
- найти размах варьирования и разбить его на 6 интервалов;
- построить полигон и гистограмму частот.

Решение

```
> restart;
```

```
> with(stats):
```

```
> x:= [9.19, 11.5, 10.7, 12.6, 13.0, 12.3,7.46,  
8.92,8.80, 11.6, 11.9, 10.9, 5.82, 8.89, 9.32, 8.30,  
8.76, 8.01, 15.5, 12.3, 9.46, 9.11, 12.1, 12.5, 9.33,  
11.0, 10.1, 9.61, 13.7, 15.0, 12.2, 13.1, 11.7, 10.4,  
11.5, 9.02, 9.23, 7.16, 12.0, 10.6, 6.39,6.97,
```

```
9.03, 6.84, 8.29, 10.5, 11.7, 7.05, 12.1,9.53];
```

```
x:= [9.19, 11.5, 10.7, 12.6, 13.0, 12.3, 7.46, 8.92, 8.80, 11.6, 11.9, 10.9, 5.82, 8.89, 9.32,  
8.30, 8.76, 8.01, 15.5, 12.3, 9.46, 9.11, 12.1, 12.5, 9.33, 11.0, 10.1, 9.61, 13.7, 15.0,  
12.2, 13.1, 11.7, 10.4, 11.5, 9.02, 9.23, 7.16, 12.0, 10.6, 6.39, 6.97, 9.03, 6.84, 8.29,  
10.5, 11.7, 7.05, 12.1, 9.53]
```

Сортируем варианты в порядке возрастания: (**transform** – группировка (преобразование) выборочных данных.), т.е. записываем вариационный ряд

```
> transform[statsort] (x) ;
```

```
[5.82, 6.39, 6.84, 6.97, 7.05, 7.16, 7.46, 8.01, 8.29, 8.30, 8.76, 8.80, 8.89, 8.92, 9.02,  
9.03, 9.11, 9.19, 9.23, 9.32, 9.33, 9.46, 9.53, 9.61, 10.1, 10.4, 10.5, 10.6, 10.7, 10.9,  
11.0, 11.5, 11.5, 11.6, 11.7, 11.7, 11.9, 12.0, 12.1, 12.1, 12.2, 12.3, 12.3, 12.5, 12.6,  
13.0, 13.1, 13.7, 15.0, 15.5]
```

Принимаем $k = 6$. Находим размах выборки:

```
> 15.5-5.82;
```

```
9.68
```

Оцениваем шаг:

```
> (%) / 6;
```

```
1.613333333
```

Округляем (**только в большую сторону**) и принимаем $h = 1,7$.
Оцениваем

крайнее левое значение первого интервала:

```
> 5.82-(1.7*6-9.68)/2;
```

5.560000000

Округляем до 5,6 и находим интервальный вариационный ряд:

```
> transform[tallyinto] (x, [5.6..7.3, 7.3..9, 9..10.7,  
10.7..12.4, 12.4..14.1, 14.1..15.8]);
```

```
[Weight(5.6 .. 7.3, 6), Weight(7.3 .. 9, 8), Weight(9 .. 10.7, 14),  
Weight(10.7 .. 12.4, 15), Weight(12.4 .. 14.1, 5), Weight(14.1 .. 15.8, 2)]
```

```
> > X:=transform[statsort](%);
```

```
X := [Weight(5.6 .. 7.3, 6), Weight(7.3 .. 9, 8), Weight(9 .. 10.7, 14),  
Weight(10.7 .. 12.4, 15), Weight(12.4 .. 14.1, 5), Weight(14.1 .. 15.8, 2)]
```

Составляем дискретный вариационный ряд:

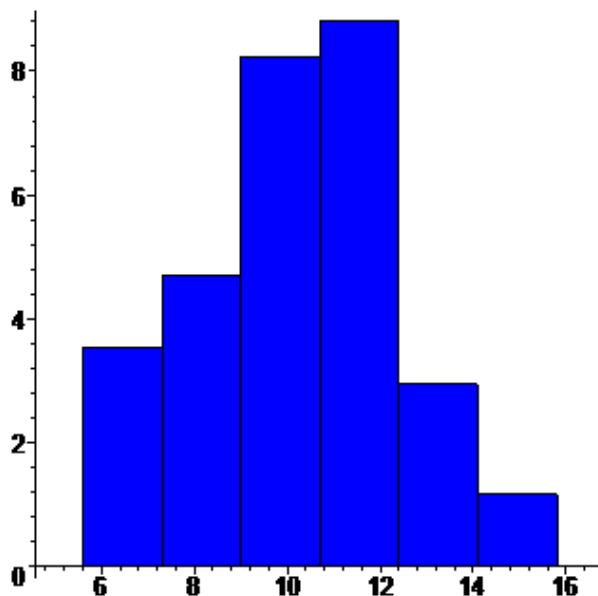
```
> Digits:=4;
```

```
> Y:=transform[classmark](X) ;
```

```
Y := [Weight(6.450, 6), Weight(8.150, 8), Weight(9.850, 14), Weight(11.55, 15),  
Weight(13.25, 5), Weight(14.95, 2)]
```

По интервальному вариационному ряду строится гистограмма частот

```
> with(stats[statplots]): histogram(X,colour=blue);
```



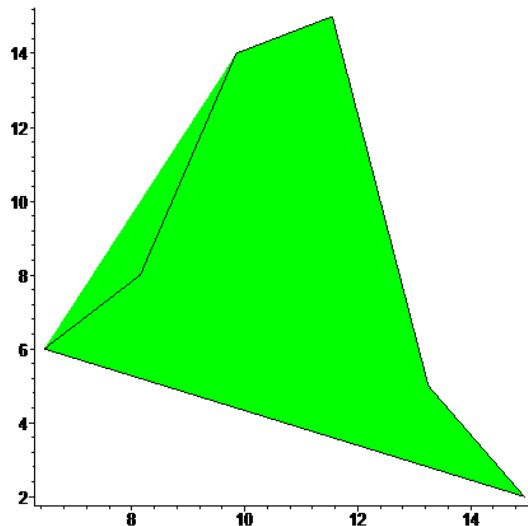
По дискретному вариационному ряду строится полигон частот:

```
> with(plottools):
```

```
>l:=polygon([[6.45,6],[8.15,8],[9.85,14],[11.55,15],[13.25,5],[14.95,2]], color = green);
```

```
l:=POLYGONS(
  [[6.45, 6.], [8.15, 8.], [9.85, 14.], [11.55, 15.], [13.25, 5.], [14.95, 2.]],
  COLOUR(RGB, 0.75294118, 0.75294118, 0.75294118))
```

```
> plots[display] (l) ;
```



Пример:

Вычислить объем выборки, выборочную среднюю и среднеквадратическое отклонение

x_i	-4.5	-2.5	-0.5	1.5	3.5	5.5	7.5
n_i	11	24	27	34	25	19	10

Решение

Вводим заданное статистическое распределение

```
> restart;
> with(stats):
> W:=[Weight(-4.5,11),Weight(-2.5,24),Weight(-0.5,27),Weight(1.5,34),Weight(3.5,25),Weight(5.5,19),Weight(7.5,10)];
>
W := [Weight(-4.5, 11), Weight(-2.5, 24), Weight(-0.5, 27), Weight(1.5, 34),
      Weight(3.5, 25), Weight(5.5, 19), Weight(7.5, 10)]
```

В подпакете describe вычисляем объем выборки, выборочную среднюю, выборочное среднее квадратическое отклонение:

> **N:=describe[count](W);**

N := 150

> **a:=describe[mean](W);**

a := 1.300000000

> **q:=describe[standarddeviation](W);**

q := 3.312602199

Задания для самостоятельного решения.

В результате эксперимента получены данные, записанные в виде статистического ряда.

Требуется:

- а) записать значения результатов эксперимента в виде вариационного ряда;
- б) найти размах варьирования и разбить его на 6 интервалов;
- в) построить полигон и гистограмму частот.

1)

17,1	21,4	15,9	19,1	22,4	20,7	17,9	18,6	21,8	16,1
19,1	20,5	14,2	16,9	17,8	18,1	19,1	15,8	18,8	17,2
16,2	17,3	22,5	19,9	21,1	15,1	17,7	19,8	14,9	20,5
17,5	19,2	18,5	15,7	14,0	18,6	21,2	16,8	19,3	17,8
18,8	14,3	17,1	19,5	16,3	20,3	17,9	23,0	17,2	15,2
15,6	17,4	21,3	22,1	20,0	14,5	19,3	18,4	16,7	18,2
16,4	18,7	14,3	18,2	19,1	15,3	21,5	17,2	22,6	20,4
22,8	17,5	20,2	15,5	21,6	18,1	20,5	14,0	18,9	16,5
20,8	16,6	18,3	21,7	17,4	23,0	21,1	19,8	15,4	18,1
18,9	14,7	19,5	20,9	15,8	20,2	21,8	18,2	21,2	20,1

2)

16,8	17,9	21,4	14,1	19,1	18,1	15,1	18,2	20,3	16,7
19,5	18,5	22,5	18,4	16,2	18,3	19,1	21,4	14,5	16,1
21,5	14,9	18,6	20,4	15,2	18,5	17,1	22,4	20,8	19,8
17,2	19,7	16,3	18,7	14,4	18,8	19,5	21,6	15,3	17,3
22,8	17,4	22,2	16,5	21,7	15,4	21,3	14,3	20,5	16,4
20,6	15,5	19,4	17,5	20,9	23,0	18,9	15,9	18,2	20,7
17,9	21,8	14,2	21,2	16,1	18,4	17,5	19,3	22,7	19,6
22,1	17,6	16,7	20,4	15,7	18,1	16,6	18,3	15,5	17,7
19,2	14,8	19,7	17,7	16,5	17,8	18,5	14,0	21,9	16,9
15,8	20,8	17,1	20,1	22,6	18,9	15,6	21,1	20,2	15,1

3)

189	207	213	208	186	210	198	219	231	227
202	211	220	236	227	220	210	183	213	190
197	227	187	226	213	191	209	196	202	235
211	214	220	195	182	228	202	207	192	226
193	203	232	202	215	195	220	233	214	185
234	215	196	220	203	236	225	221	193	215
204	184	217	193	216	205	197	203	229	204
225	216	233	223	208	204	207	182	216	191
210	190	207	205	232	222	198	217	211	201
185	217	225	201	208	211	189	205	207	199

4).

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

5).

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

4,8	6,1	3,6	9,5	8,4	2,4	6,2	7,3	5,7	0,9
7,4	8,5	5,8	1,1	5,9	4,9	3,7	9,6	2,6	6,1