Пример решения самостоятельной работы №1 по математической статистике

Вариант $* * *$ дата ___.__.2014 преподаватель: В.И. Глебов

1. По данной выборке $\{-12; 8; 3; 1; 11; 1; 11; 11; -12; 1;\}$ построить вариационный и статистический ряды, а также эмпирическую функцию распределения.

Ответ:

Вариационный ряд:

Статистический ряд:

i	1	2	3	4	5	6	7	8	9	10
x_i										
n_i										

Эмпирическая функция распределения:

i	0	1	2	3	4	5	6	7	8	9	10	
диапазон x												
$\sum n_j$												
F_i^*												

2. Для данного интервального статистического ряда заполнить таблицу относительных частот группированной выборки, построить эмпирическую функцию распределения (кумуляту) и гистограмму приведённых относительных частот ν_i/Δ_i.

i	$x_{i-1} \leq x < x_i$	n_i	Δ_i	z_i	$\nu_i = n_i/n$	ν_i/Δ_i	$\sum_{j=1}^{i} \nu_j$	$F^*(x)$
1	5 – 7	4						
2	7 – 9	9						
3	9 – 11	7						
4	11 – 13	7						
5	13 – 15	6						
6	15 – 17	2						
7	17 – 19	5						

1. По данной выборке $\{-12; 8; 3; 1; 11; 1; 11; 11; -12; 1;\}$ построить вариационный и статистический ряды, а также эмпирическую функцию распределения.

Решение.

1 задание.

Вариационный ряд получается из выборки путём её упорядочивания в порядке возрастания элементов (ранжирования).

1 способ.

Наименьшими в выборке являются первый и 9-й элементы (-12). Они занимают самые первые позиции вариационного ряда.

$$\left\{\boxed{-12}; 8; 3; 1; 11; 1; 11; 11; \boxed{-12}; 1;\right\} \Rightarrow$$

−12	−12								

Следующие позиции занимают наименьшие среди оставшихся (1).

$$\left\{\boxed{-12}; 8; 3; \boxed{1}; 11; \boxed{1}; 11; 11; \boxed{-12}; \boxed{1};\right\} \Rightarrow$$

−12	−12	1	1	1					

Затем находим 3

$$\left\{\boxed{-12}; 8; \boxed{3}; \boxed{1}; 11; \boxed{1}; 11; 11; \boxed{-12}; \boxed{1};\right\} \Rightarrow$$

−12	−12	1	1	1	3				

Затем 8

$$\left\{\boxed{-12}; \boxed{8}; \boxed{3}; \boxed{1}; 11; \boxed{1}; 11; 11; \boxed{-12}; \boxed{1};\right\} \Rightarrow$$

−12	−12	1	1	1	3	8			

И, наконец, остаются три совпадающих элемента (11)

$$\left\{\boxed{-12}; \boxed{8}; \boxed{3}; \boxed{1}; \boxed{11}; \boxed{1}; \boxed{11}; \boxed{11}; \boxed{-12}; \boxed{1};\right\} \Rightarrow$$

−12	−12	1	1	1	3	8	11	11	11

В результате построен вариационный ряд.

2 способ

Записываем выборку в первые две строки таблицы, а третью строку оставляем для записи рангов.

Начинаем с присвоения начальных значений рангов наименьшим элементам выборки

i	1	2	3	4	5	6	7	8	9	10
x_i	-12	8	3	1	11	1	11	11	-12	1
r_i	1								2	

Затем — следующим

i	1	2	3	4	5	6	7	8	9	10
x_i	-12	8	3	1	11	1	11	11	-12	1
r_i	1			3		4			2	5

$\Rightarrow$

i	1	2	3	4	5	6	7	8	9	10
x_i	-12	8	3	1	11	1	11	11	-12	1
r_i	1	7	6	3		4			2	5

$\Rightarrow$

i	1	2	3	4	5	6	7	8	9	10
x_i	-12	8	3	1	11	1	11	11	-12	1
r_i	1	7	6	3	8	4	9	10	2	5

Теперь остаётся упорядочить таблицу в порядке возрастания ранга

i	1	9	4	6	10	3	2	5	7	8
x_i	-12	-12	1	1	1	3	8	11	11	11
r_i	1	2	3	4	5	6	7	8	9	10

Вариационный ряд расположен во второй строке таблицы

Ответ: Вариационный ряд:

−12	−12	1	1	1	3	8	11	11	11

2 задание.

Статистический ряд получается из вариационного, если подсчитать количество повторений n_i (частоту) для каждого элемента вариационного ряда x_i и поставить это количество в соответствие значению этого элемента (варианте)

x_i	$\underbrace{-12,-12}$	$\underbrace{1,1,1}$	3	8	$\underbrace{11,11,11}$
n_i	2	3	1	1	3

$\Rightarrow$

i	1	2	3	4	5
x_i	−12	1	3	8	11
n_i	2	3	1	1	3

Ответ: Статистический ряд:

i	1	2	3	4	5
x_i	−12	1	3	8	11
n_i	2	3	1	1	3

3 задание.

Эмпирическая функция распределения определяется как доля тех элементов выборки, значения которых строго меньше аргумента функции.

Статистический ряд в условиях данной задачи представляет собой статистический аналог ряда распределения дискретной случайной величины. По этой причине эмпирическая функция распределения имеет вид, аналогичный функции распределения дискретной случайной величины, т.е. вид кусочно-постоянной функции.

Точки разрыва соответствуют элементам статистического ряда (вариантам). Во всех точках разрыва функция непрерывна слева.

Делим всю числовую ось от $-\infty$ к $+\infty$ на ряд промежутков постоянства функции распределения.

Первый промежуток $x \in (-\infty;-12]$ не содержит элементов статистического ряда, удовлетворяющих условию $x_i < x$. Поэтому доля таких элементов равна нулю $\Rightarrow F_i(x) = 0$.

i	0	1	2	3	4	5
диапазон x	$(-\infty;-12]$					
$\sum n_i$	0					

Второй промежуток располагается между первой и второй вариантами $x \in (-12;1]$. В этом диапазоне два элемента выборки, представленные первыми двумя элементами вариационного ряда или первым элементом статистического ряда, удовлетворяют условию $x_i < x \Rightarrow$

i	0	1	2	3	4	5
диапазон x	$(-\infty;-12]$	$(-12;1]$				
$\sum n_i$	0	2				

Третий промежуток располагается между второй и третьей вариантами $x \in (1;3]$. В этом диапазоне кроме первого элемента статистического ряда, условию $x_i < x$ удовлетворяет таже второй элемент статистического ряда $x_2 = 1 < x$. Поэтому к двум элементам выборки, соответствующим первой варианте, мы должны прибавить и все элементы выборки, соответствующие второй варианте. Всего для третьего промежутка будет $n_1 + n_2 = 2 + 3 = 5$ элементов.

i	0	1	2	3	4	5
диапазон x	$(-\infty;-12]$	$(-12;1]$	$(1;3]$			
$\sum n_i$	0	2	5			

$\Rightarrow$

Продолжая эти рассуждения, получаем следующую цепочку действий

i	0	1	2	3	4	5
диапазон x	$(-\infty;-12]$	$(-12;1]$	$(1;3]$	$(3;8]$		
$\sum n_i$	0	2	5	6		

$\Rightarrow$

i	0	1	2	3	4	5
диапазон x	$(-\infty;-12]$	$(-12;1]$	$(1;3]$	$(3;8]$	$(8;11]$	
$\sum n_i$	0	2	5	6	7	

$\Rightarrow$

i	0	1	2	3	4	5
диапазон x	$(-\infty;-12]$	$(-12;1]$	$(1;3]$	$(3;8]$	$(8;11]$	$(11;\infty)$
$\sum n_i$	0	2	5	6	7	10

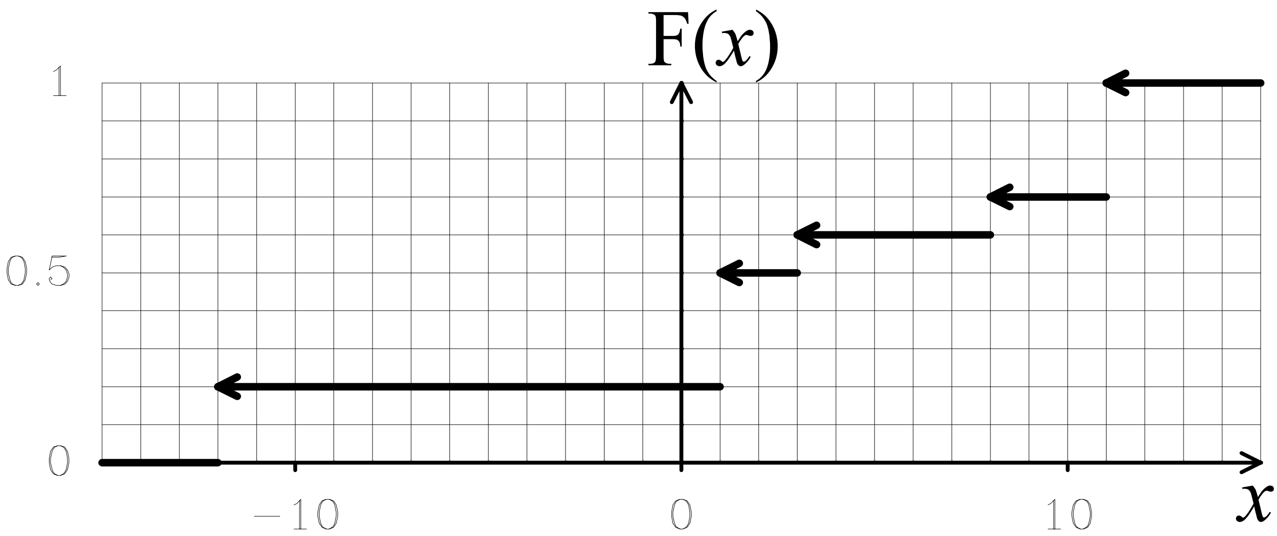
Объём выборки $n = \sum_{j=1}^{5} n_j = 10$.

Для нахождения эмпирической функции распределения надо найти соотвествующие доли объёма выборки, т.е. для каждого промежутка найти отношение накопленных частот к объёму выборки $\frac{\sum_{j=1}^{i} n_j}{n}$

В результате получаем эмпирическую функцию распределения:

i	0	1	2	3	4	5
диапазон x	$(-\infty;-12]$	$(-12;1]$	$(1;3]$	$(3;8]$	$(8;11]$	$(11;\infty)$
$\sum n_i$	0	2	5	6	7	10
F_i	0	$\frac{2}{10}$	$\frac{5}{10}$	$\frac{6}{10}$	$\frac{7}{10}$	$\frac{10}{10}$

Осталось построить график, стараясь соблюсти верный масштаб



2. Для данного интервального статистического ряда заполнить таблицу относительных частот группированной выборки, построить эмпирическую функцию распределения (кумуляту) и гистограмму приведённых относительных частот ν_i/Δ_i.

1	2	3	4	5	6	7	8	9
i	$x_{i-1} \le x < x_i$	n_i	Δ_i	z_i	$\nu_i = n_i/n$	ν_i/Δ_i	$\sum_{j=1}^{i} \nu_j$	$F^*(x)$
1	5 – 7	4						
2	7 – 9	9						
3	9 – 11	7						
4	11 – 13	7						
5	13 – 15	6						
6	15 – 17	2						
7	17 – 19	5						

Решение.

4-я и 5-я колонки содержат ширину интервала Δ_i и "псевдоварианту" z_i, которая представляет данный интервал в разного рода оценочных расчётах. Обычно в качестве z_i выбирают середину интервала $\frac{x_{i-1} + x_i}{2}$. Так и сделаем.

1	2	3	4	5	6	7	8	9
i	$x_{i-1} \le x < x_i$	n_i	Δ_i	z_i	$\nu_i = n_i/n$	ν_i/Δ_i	$\sum_{j=1}^{i} \nu_j$	$F^*(x)$
1	5 – 7	4	2	6				
2	7 – 9	9	2	8				
3	9 – 11	7	2	10				
4	11 – 13	7	2	12				
5	13 – 15	6	2	14				
6	15 – 17	2	2	16				
7	17 – 19	5	2	18				

Находим объём выборки, складывая все частоты n_i. В нижней строке третьей колонки записываем результат

$$n = \sum_{i=1}^{i_{\max}} n_i = \sum_{i=1}^{7} n_i = 4 + 9 + 7 + 7 + 6 + 2 + 5 = 40.$$

В 6-ю колонку записываем относительные частоты (частости) $\nu_i = n_i/n$, каждая из которых означает долю объёма выбоки, приходяшуюся на данный интервал. В нижней строке этой колонки записываем сумму дробей $\sum_{i=1}^{i_{\max}} \nu_i$. Должна получиться 1. Если не получилась, ищем ошибку (либо неверно найден объём выборки n, либо неверно записаны какие-то из дробей $\nu_i = n_i/n$).

1	2	3	4	5	6	7	8	9
i	$x_{i-1} \le x < x_i$	n_i	Δ_i	z_i	$\nu_i = n_i/n$	ν_i/Δ_i	$\sum_{j=1}^{i} \nu_j$	$F^*(x)$
1	5 – 7	4	2	6	4/40			
2	7 – 9	9	2	8	9/40			
3	9 – 11	7	2	10	7/40			
4	11 – 13	7	2	12	7/40			
5	13 – 15	6	2	14	6/40			
6	15 – 17	2	2	16	2/40			
7	17 – 19	5	2	18	5/40			
		40			1			

Теперь всё готово к заполнению 7-й и 8-й колонок (приведённых относительных частот ν_i/Δ_i и накопленных частот $\sum_{j=1}^{i} \nu_j$).

7-ю колнку получаем из 6-й путём деления на Δ_i (в нашем примере все $\Delta_i = 2$), а $\sum_{j=1}^{i} \nu_j$ получаем путём добавления к уже имеющейся сумме величины ν_i из 6-й колонки.Добавляем в таблицу строку, соответствующую "нулевому" интервалу, считая, что $\sum_{j=1}^{-1} \nu_j = \frac{0}{40}$

1	2	3	4	5	6	7	8	9
i	$x_{i-1} \le x < x_i$	n_i	Δ_i	z_i	$\nu_i = n_i/n$	ν_i/Δ_i	$\sum_{j=1}^{i} \nu_j$	$F^*(x)$
0	$-\infty$ – 5	0			0/40	0/80	0/40	
1	5 – 7	4	2	6	4/40	4/80	$\frac{0+4}{40} = \frac{4}{40}$	
2	7 – 9	9	2	8	9/40	9/80	$\frac{4+9}{40} = \frac{13}{40}$	
3	9 – 11	7	2	10	7/40	7/80	$\frac{13+7}{40} = \frac{20}{40}$	
4	11 – 13	7	2	12	7/40	7/80	$\frac{27}{40}$	
5	13 – 15	6	2	14	6/40	6/80	$\frac{33}{40}$	
6	15 – 17	2	2	16	2/40	2/80	$\frac{35}{40}$	
7	17 – 19	5	2	18	5/40	5/80	$\frac{40}{40} = 1$	
		40			1			

Найденные в 8-й колонке числа являются значениями эмпирической функции распределения $F^*(x_i)$ в точках x_i, соответствующих правым границам соответствующих (i-ых) интервалов.

Аналитическое выражение для $F^*(x)$ для i-го интервала легко найти, если предположить, что зависимость будет линейной. (Это предположение основано на том факте, что отсутствует информация о распределении вероятностей внутри интервалов и мы ограничиваемся простейшим предположением — что распределение внутри интервала равномерное). Поэтому записываем уравнение прямой, проходящей через две заданные точки $(x_{i-1}; F^*(x_{i-1}))$ и $(x_i; F^*(x_i))$. Вспомним школу, первый курс и запишем:

$$F^*(x) = F^*(x_{i-1}) + \frac{x - x_{i-1}}{x_i - x_{i-1}}(F^*(x_i) - F^*(x_{i-1})) = \sum_{j=1}^{i-1} \nu_j + \frac{x - x_{i-1}}{\Delta_i} \cdot \nu_i = \sum_{j=1}^{i-1} \nu_j + \frac{\nu_i}{\Delta_i} \cdot (x - x_{i-1})$$

1	2	3	4	5	6	7	8	9
i	$x_{i-1} \le x < x_i$	n_i	Δ_i	z_i	$\nu_i = n_i/n$	ν_i/Δ_i	$\sum_{j=1}^{i} \nu_j$	$F^*(x)$
0	$-\infty - 5$	0			0/40	0/80	$\frac{0}{40} \searrow$	
1	$5 - 7$	4	2	6	4/40	4/80	$\frac{4}{40} \searrow$	$\searrow \frac{0}{40} + \frac{4}{80} \cdot (x-5)$
2	$7 - 9$	9	2	8	9/40	9/80	$\frac{13}{40} \searrow$	$\searrow \frac{4}{40} + \frac{9}{80} \cdot (x-7)$
3	$9 - 11$	7	2	10	7/40	7/80	$\frac{20}{40} \searrow$	$\searrow \frac{13}{40} + \frac{7}{80} \cdot (x-9)$
4	$11 - 13$	7	2	12	7/40	7/80	$\frac{27}{40} \searrow$	$\searrow \frac{20}{40} + \frac{7}{80} \cdot (x-11)$
5	$13 - 15$	6	2	14	6/40	6/80	$\frac{33}{40} \searrow$	$\searrow \frac{27}{40} + \frac{6}{80} \cdot (x-13)$
6	$15 - 17$	2	2	16	2/40	2/80	$\frac{35}{40} \searrow$	$\searrow \frac{33}{40} + \frac{2}{80} \cdot (x-15)$
7	$17 - 19$	5	2	18	5/40	5/80	$\frac{40}{40} = 1$	$\searrow \frac{35}{40} + \frac{5}{80} \cdot (x-17)$
$\sum$		40			1			

Добавляем значения за пределами диапазонов возможных значений

1	2	3	4	5	6	7	8	9
i	$x_{i-1} \le x < x_i$	n_i	Δ_i	z_i	$\nu_i = n_i/n$	ν_i/Δ_i	$\sum_{j=1}^{i} \nu_j$	$F^*(x)$
0	$-\infty - 5$	0			0	0	0	0
1	$5 - 7$	4	2	6	4/40	4/80	$\frac{4}{40}$	$0 + \frac{4}{80} \cdot (x-5)$
2	$7 - 9$	9	2	8	9/40	9/80	$\frac{13}{40}$	$\frac{4}{40} + \frac{9}{80} \cdot (x-7)$
3	$9 - 11$	7	2	10	7/40	7/80	$\frac{20}{40}$	$\frac{13}{40} + \frac{7}{80} \cdot (x-9)$
4	$11 - 13$	7	2	12	7/40	7/80	$\frac{27}{40}$	$\frac{20}{40} + \frac{7}{80} \cdot (x-11)$
5	$13 - 15$	6	2	14	6/40	6/80	$\frac{33}{40}$	$\frac{27}{40} + \frac{6}{80} \cdot (x-13)$
6	$15 - 17$	2	2	16	2/40	2/80	$\frac{35}{40}$	$\frac{33}{40} + \frac{2}{80} \cdot (x-15)$
7	$17 - 19$	5	2	18	5/40	5/80	$\frac{40}{40} = 1$	$\frac{35}{40} + \frac{5}{80} \cdot (x-17)$
8	$19 - \infty$	0			0	0	1	1
$\sum$		40			1			

Теперь можно строить графики.

Начинать построение следует с выбора подходящего масштаба и разметки осей.

Если на осях имеются шкалы (серия засечек), следует использовать их, поставив соответствующие метки (числа). Метки надо ставить равномерно, вне связи с наносимыми впоследствие точками.

Масштаб вдоль оси ординат выбирается таким, чтобы он занимал не менее половины высоты и был бы удобен для нанесения точек и проведения линий.

График эмпирической плотности является гистограммой и строится на основе значений ν_i/Δ_i (из 7-ой колонки). Максимальное значение в этой колонке 9/80. Пример разумной разметки оси ординат можно увидеть на рисунках.

Графики без разметки осей считаются непостроеными!

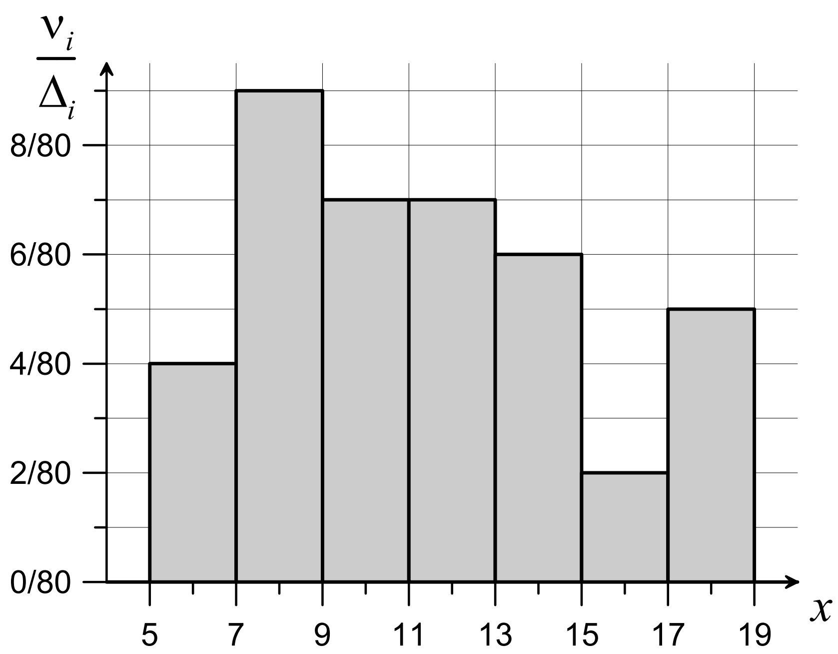


График эмпирической функции распределения (кумуляты) строим на основе значений из 8-ой колонки ($\sum_{j=1}^{i} \nu_j$).

На оси ординат выбираем для единицы ($1 = n/n$) количество делений, равное n. Или такое, при котором легко откладывать числители дробей в 8-ой колонке. Сначала ставим точки $(5; 0/40)$, затем $(7; 4/40)$, затем $(9; 13/40)$, затем …, затем $(19; 40/40)$.

Проверяем правильность нанесения точек и только после этого соединяем точки отрезками прямых, проводим влево горизонтальный луч от точки $(5; 0/40)$, и вправо от точки $(19; 40/40)$. Любуемся работой и ВСЁ!