Практическое занятие 2: ***Дискретный вариационный ряд. Полигон частот и эмпирическая функция распределения***

Цель занятия:

- расчёт выборочной средней;
- приобретение навыков составления вариационного ряда и построения ***полигона частот***;
- приобретение навыков расчёта относительных частот и построения ***эмпирической функции распределения***.

Задание 1

Варианты задания 1

Задание 2

Варианты задания 2

Дискретный вариационный ряд – это упорядоченное по возрастанию *(как правило)* множество *вариант* $x_1, x_2, \ldots, x_k$ (значений величины X) и соответствующих им *частот* либо ***относительных частот***.

Частоты *выборочной совокупности* обозначают через $n_1, n_2, \ldots, n_k$, частоты *генеральной совокупности*– через $N_1, N_2, \ldots, N_k$.

Относительные частоты рассчитываются по формулам:

$w_1=n_1/n,\ w_2=n_2/n,\ w_3=n_3/n,\ \ldots,\ w_k=n_k/n.$

где $n_1+n_2+\ldots+n_k$ *объем* выборки, при этом, сумма всех относительных частот: $w_1+w_2+\ldots+w_k=1$.

Аналогично для совокупности генеральной:

$W_1=N_1/N,\ W_2=N_2/N,\ W_3=N_3/N,\ \ldots,\ W_k=N_k/N$,

где $N=N_1+N_2+\ldots+N_k$ – её *объем*, и, очевидно: $W_1+W_2+\ldots+W_k=1$.

Задание 1

По результатам **выборочного исследования** рабочих цеха были установлены их квалификационные разряды: 4, 5, 6, 4, 4, 2, 3, 5, 4, 4, 5, 2, 3, 3, 4, 5, 5, 2, 3, 6, 5, 4, 6, 4, 3.

Требуется:

- составить вариационный ряд и построить ***полигон частот***;
- найти относительные частоты и построить ***эмпирическую функцию распределения***.

Методика выполнения

В условии прямо сказано о том, что имеется *выборка* из генеральной совокупности (всех рабочих цеха), и первое, что логично сделать – подсчитать её *объем*, т.е. количество рабочих. В данном случае n=25.

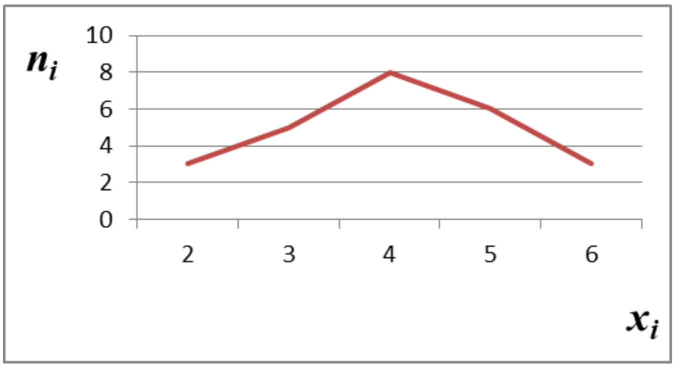
Квалификационные разряды – есть величина *дискретная*, и поэтому предстоит составить дискретный вариационный ряд.

x_i	n_i
2	3
3	5
4	8
5	6
6	3
Σ	25

Полученные результаты позволяют достаточно точно судить об уровне квалификации всего цеха.

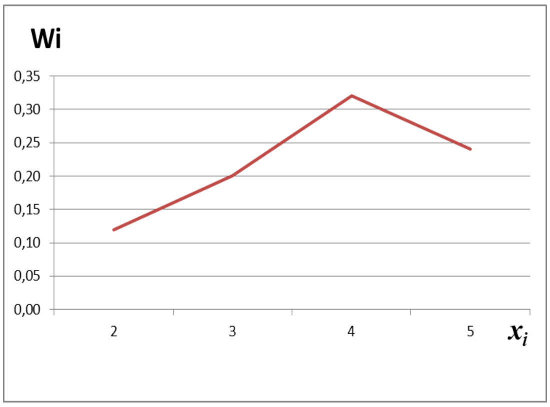
Построенный вариационный ряд также называют ***статистическим распределением выборки***, причём, этот термин применим не только для дискретного, но и для **интервального ряда**, о котором позже.

Построим полигон частот. Это статистический аналог **многоугольника распределения дискретной случайной величины**. Полигон частот – это *ломаная*, соединяющая соседние точки ($\boldsymbol{x_i}$, $\boldsymbol{n_i}$):



Иногда требуется построить ***полигон относительных частот*** - ломаную, соединяющую соседние точки ($\boldsymbol{x_i}$, $\boldsymbol{w_i}$).. Но такое задание больше характерно для **интервального вариационного ряда**.

x_i	W_i	N_i
2	0,12	3
3	0,20	5
4	0,32	8
5	0,24	6
6	0,12	3
Σ	1	25



А теперь посмотрим на относительные частоты и задумаемся, на что они похожи? …Правильно, на вероятности. Так, например, можно сказать, что w_3=0,32 – есть *примерная* вероятность того, что наугад выбранный рабочий цеха будет иметь 4 разряд. «Примерная» – по той причине, что перед нами выборка.

А вот если учесть ВСЕХ рабочих цеха (всю генеральную совокупность), то рассчитанные относительные частоты W_1, W_2, W_3, $W_{4/}$, W_5 – и есть в точности эти вероятности.

Построим ***эмпирическую функцию распределения F*(x)***. Это статистический аналог **функции распределения** из теории вероятностей. Данная функция определяется, как отношение:

$\boldsymbol{F^*(x) = n_1/n}$ **,** , где $\boldsymbol{n_1}$ – количество вариант СТРОГО МЕНЬШИХ, чем x, при этом «икс» «пробегает» все значения от «минус» до «плюс» бесконечности.

Очевидно, что на интервале $x \in (-\infty, 2)$ $F^*(x)=0$, и, кроме того, функция равна нулю ещё и в точке $F^*(2)=0$. Почему? Потому, что значение $F^*(2)$ определяет количество вариант, которые СТРОГО меньше двух, а это количество равно нулю.

На промежутке $\boldsymbol{x \in (2, 3)}$ $\boldsymbol{F^*(x) = n_x/n = n_1/n = w_1}$ =0,12 – и опять обратите внимание, что значение $F^*(3)$ не учитывает рабочих 3-го разряда, т.к. речь идёт о вариантах, которые СТРОГО меньше трёх.

На промежутке $\boldsymbol{x \in (3, 4)}$ $\boldsymbol{F^*(x) = n_x/n = (n_1+n_2)/n = w_1+w_2}$ =0,12+0,2=0,32 и далее процесс продолжается по принципу накопления частот:

– если $4 < x \le 5$, то $F^*(x)=$ $\boldsymbol{(n_1+n_2+n_3)/n = w_1+w_2+w_3}$=0,12+0,2+0,32 =0,64;

– если $5 < x \le 6$, то $F^*(x)=$ $\boldsymbol{= w_1+w_2+w_3+w_4}$=0,12+0,2+0,32+0,24 =0,88;

– наконец, если $6 < x \le +\infty$, то $F^*(x)=$ $\boldsymbol{= w_1+w_2+w_3+w_4 +w_5}$=0,12+0,2+0,32+0,24 +0,12 =1 - и в самом деле, для ЛЮБОГО «икс» из интервала $(6, +\infty)$ ВСЕ частоты расположены СТРОГО левее этого «икс».

Накопленные относительные частоты удобно записывать в отдельный столбец таблицы, при этом алгоритм вычислений очень прост: сначала сносим слева 1-е значение, а каждое следующее получаем как сумму предыдущего и относительной частоты из текущего левого столбца:

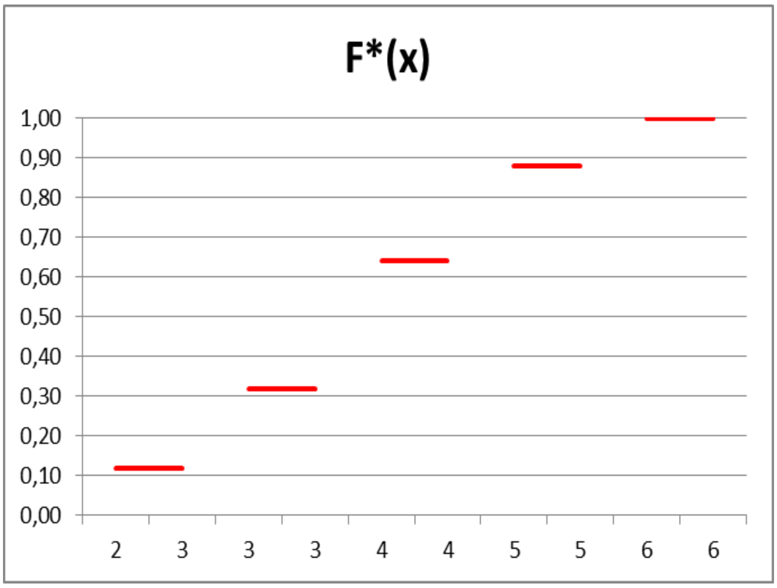
x_i	N_i	W_i	$W_н$
2	3	0,12	0,12
3	5	0,20	0,32
4	8	0,32	0,64
5	6	0,24	0,88
6	3	0,12	1,00
Σ	25	1	

=D11+C12

Саму функцию принято записывать в кусочном виде:

$$F^*(x) = \begin{cases} 0,\ \textit{если}\ \ x \le 2 \\ 0{,}12,\ \textit{если}\ \ 2 < x \le 3 \\ 0{,}32,\ \textit{если}\ \ 3 < x \le 4 \\ 0{,}64,\ \textit{если}\ \ 4 < x \le 5 \\ 0{,}88,\ \textit{если}\ \ 5 < x \le 6 \\ 1,\ \textit{если}\ \ x > 6 \end{cases}$$

а её график представляет собой ступенчатую фигуру:



Эмпирическая функция распределения ***не убывает*** и принимает значения из промежутка ***0 ≤ F*(x) ≤ 1***.

Эмпирическая функция распределения F*(x) строится по выборке и приближает ***теоретическую функцию распределения F(x)***. Легко догадаться, что последняя образуется на основании исследования всей генеральной совокупности, но если рабочих в цехе ещё пересчитать можно, то звёзды на небе – уже вряд ли. Вот поэтому и важна именно эмпирическая функция, и ещё важнее, чтобы выборка была *репрезентативна*, дабы приближение было хорошим.

Варианты задания 1

№ вар	*Квалификационные разряды выборки*														
1	**5**	**5**	**5**	**5**	**4**	**3**	**4**	**3**	**3**	**3**	**4**	**6**	**6**	**2**	**5**
2	**5**	**2**	**2**	**5**	**5**	**3**	**3**	**3**	**3**	**3**	**2**	**6**	**2**	**3**	**4**
3	**2**	**4**	**3**	**3**	**3**	**6**	**5**	**5**	**4**	**6**	**3**	**2**	**6**	**6**	**2**
4	**5**	**5**	**6**	**2**	**6**	**6**	**4**	**2**	**4**	**4**	**2**	**3**	**3**	**5**	**6**
5	**4**	**5**	**2**	**5**	**4**	**6**	**6**	**4**	**3**	**3**	**5**	**5**	**6**	**2**	**6**
6	**4**	**2**	**4**	**6**	**5**	**5**	**5**	**5**	**5**	**3**	**3**	**5**	**5**	**4**	**4**
7	**3**	**5**	**3**	**6**	**3**	**4**	**4**	**3**	**2**	**2**	**2**	**5**	**6**	**6**	**4**
8	**2**	**4**	**6**	**4**	**4**	**3**	**4**	**3**	**2**	**6**	**2**	**2**	**3**	**4**	**4**
9	**5**	**4**	**6**	**4**	**3**	**5**	**3**	**5**	**2**	**6**	**2**	**2**	**6**	**6**	**6**
10	**6**	**5**	**3**	**4**	**3**	**4**	**2**	**3**	**3**	**5**	**2**	**4**	**4**	**5**	**3**
11	**2**	**3**	**5**	**5**	**2**	**6**	**6**	**6**	**2**	**5**	**2**	**3**	**2**	**5**	**6**
12	**6**	**5**	**6**	**5**	**5**	**2**	**3**	**4**	**5**	**5**	**4**	**2**	**6**	**3**	**6**
13	**6**	**4**	**5**	**2**	**2**	**4**	**5**	**3**	**5**	**2**	**4**	**3**	**3**	**3**	**4**
14	**5**	**3**	**6**	**5**	**5**	**2**	**2**	**2**	**6**	**5**	**5**	**5**	**2**	**3**	**3**
15	**2**	**2**	**2**	**5**	**4**	**3**	**3**	**2**	**6**	**2**	**2**	**3**	**4**	**4**	**4**
16	**2**	**3**	**4**	**2**	**2**	**4**	**3**	**6**	**5**	**3**	**6**	**2**	**2**	**4**	**6**
17	**4**	**6**	**2**	**3**	**2**	**6**	**2**	**5**	**2**	**2**	**4**	**3**	**3**	**5**	**4**
18	**4**	**5**	**5**	**4**	**2**	**2**	**3**	**3**	**4**	**6**	**6**	**2**	**6**	**3**	**4**
19	**3**	**5**	**6**	**6**	**4**	**6**	**4**	**2**	**3**	**2**	**3**	**3**	**3**	**3**	**2**
20	**4**	**6**	**4**	**5**	**2**	**4**	**3**	**6**	**3**	**5**	**2**	**6**	**6**	**2**	**3**
21	**3**	**2**	**2**	**3**	**5**	**3**	**4**	**4**	**3**	**4**	**4**	**2**	**6**	**4**	**5**
22	**6**	**6**	**3**	**2**	**4**	**5**	**4**	**3**	**3**	**5**	**2**	**3**	**5**	**5**	**2**
23	**4**	**6**	**5**	**6**	**3**	**5**	**3**	**2**	**4**	**2**	**3**	**6**	**2**	**6**	**2**
24	**3**	**6**	**3**	**6**	**6**	**2**	**2**	**5**	**3**	**5**	**4**	**6**	**2**	**6**	**5**
25	**6**	**3**	**6**	**5**	**4**	**6**	**4**	**6**	**4**	**6**	**4**	**4**	**6**	**3**	**2**

Задание 2

Дано статистическое распределение выборки

x_i	-2	1,5	5	7
n_i	12	8	20	10

Составить эмпирическую функцию распределения.

Методика выполнения

Заполним расчётную таблицу:

x_i	n_i	w_i	$w_н$
-2	12	0,24	0,24
1,5	8	0,16	0,4
5	20	0,4	0,8
7	10	0,2	1
Суммы:	50	1	

По результатам правого столбца составляется эмпирическая функция распределения

$$F^*(x) = \begin{cases} 0, \text{ } если \text{ } x \le -2 \\ 0{,}24, \text{ } если \text{ } -2 < x \le 1{,}5 \\ 0{,}4, \text{ } если \text{ } 1{,}5 < x \le 5 \\ 0{,}8, \text{ } если \text{ } 5 < x \le 7 \\ 1, \text{ } если \text{ } x > 7 \end{cases}$$

Варианты задания 2

№ ва	x_i				n_i			
1	2	3	6	2	17	12	17	18
2	6	8	2	-3	14	8	15	12
3	8	5	4	-1	16	14	10	18
4	8	0	7	1	16	10	12	9
5	8	8	2	8	17	15	11	18
6	-2	-1	0	8	12	9	8	9
7	8	9	9	8	10	14	13	9
8	2	5	2	0	12	17	14	12
9	4	0	7	1	8	12	13	9
10	0	5	5	9	17	13	10	14
11	-2	-2	8	7	14	17	18	8
12	-2	1	1	1	18	10	14	12
13	4	6	4	4	10	14	18	16
14	2	8	3	1	18	8	11	8
15	1	6	1	7	14	11	10	15
16	9	5	4	-1	17	14	16	11
17	0	1	3	2	10	12	18	9
18	2	4	0	-1	17	16	15	11
19	6	-2	2	4	11	8	16	9
20	2	9	0	-1	14	15	18	17
21	-2	2	7	2	13	13	9	13
22	-3	0	5	7	17	11	18	8
23	-3	-2	-3	-2	11	8	14	13
24	-3	7	-3	8	12	11	16	15
25	1	4	1	2	8	12	10	13