Практическое занятие 3: *Интервальный вариационный ряд. Гистограмма относительных частот*

Цель занятия:

- приобретение навыков составления интервального вариационного ряда распределения;
- построение гистограммы, полигона относительных частот и эмпирической функции распределения.

Предпосылкой построения *интервального вариационного ряда* (ИВР) является тот факт, что исследуемая величина принимает слишком много различных значений. Зачастую ИВР появляется в результате измерения *непрерывной* характеристики изучаемых объектов. Типично – это время, масса, размеры и другие физические характеристики.

Для изучения интервального вариационного ряда затруднительно либо невозможно применить тот же подход, что и для дискретного ряда. Это связано с тем, что ВСЕ *варианты* многих ИВР различны. И даже если встречаются совпадающие значения, например, 50 грамм и 50 грамм, то связано это с округлением, ибо полученные значения всё равно отличаются хоть какими-то микрограммами.

Поэтому для исследования ИВР используется другой подход, а именно, определяется интервал, в пределах которого варьируются значения, затем данный интервал делится на *частичные интервалы*, и по каждому интервалу подсчитываются *частоты* – количество *вариант*, которые в него попали.

Задание 1

По результатам исследования цены некоторого товара в различных торговых точках города, получены следующие данные (в некоторых денежных единицах).

Требуется составить вариационный ряд распределения, построить гистограмму, полигон относительных частот и эмпирическую функцию распределения.

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

Методика выполнения

Очевидно, что имеется выборочная совокупность *объемом n=30* наблюдений, и вопрос номер один: какой ряд составлять – дискретный или интервальный?

В таблице: среди предложенных цен есть одинаковые, но их разброс довольно велик, и поэтому здесь целесообразно провести интервальное разбиение. К тому же цены могут быть округлёнными.

Тактика действий похожа на исследование дискретного вариационного ряда. Сначала окидываем взглядом предложенные числа и определяем примерный интервал, в который вписываются эти значения. «Навскидку» все значения заключены в пределах от 5 до 11. Далее делим этот интервал на удобные подынтервалы, в данном случае напрашиваются промежутки единичной длины. Записываем их на черновик:

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

После этого находим самое маленькое число в левой колонке и самое большое значение – в правой.

$\boldsymbol{x_{min}}$ = 5,8, $\boldsymbol{x_{max}}$= 10,6 ден. ед.

Вычислим *размах вариации*: R= $\boldsymbol{x_{max}}$ - $\boldsymbol{x_{min}}$ = 10,6-5,8= 4,8 ден. ед. – длина общего интервала, в пределах которого варьируется цена.

Теперь его нужно разбить на *частичные интервалы*. Сколько интервалов рассмотреть? По умолчанию на этот счёт существует *формула Стерджеса*:

k=1+3,322 lg(n),

где lg(n) – десятичный логарифм от объёма выборки и k – оптимальное количество интервалов, при этом результат округляют до ближайшего левого целого значения.

В нашем случае получаем: k=1+3,322 lg(30) ≈ 5,9≈ 5 интервалов.

Следует отметить, что правило Стерджеса носит рекомендательный, но не обязательный характер. Нередко в условии задачи прямо сказано, на какое количество интервалов нужно проводить разбиение (на 4, 5, 6, 10 и т.д.), и тогда следует придерживаться именно этого указания.

Длины *частичных интервалов* могут быть различны, но в большинстве случаев использует *равноинтервальную группировку*:

$\boldsymbol{h=(x_{max}-x_{min})/k} = 4{,}8/5 = 0{,}96 \approx 1$ – длина частичного интервала.

В принципе, здесь можно было не округлять и использовать длину 0,96, но удобнее 1.

И коль скоро мы прибавили 0,04, то по 5 частичным интервалам у нас получается «перебор»: 0,04*5 = 0,2.

Поэтому от самой малой варианты x_{min} = 5,8 отмеряем влево 0,1 *(половину «перебора»)* и к значению 5,7 начинаем прибавлять по h=1, получая тем самым частичные интервалы.

При этом сразу рассчитываем их середины x_i (например, x_1 = (5,7+6,7)/2 = 6.2) – они требуются почти во всех тематических задачах:

Интервалы		xi
5,7	6,7	6,2
6,7	7,7	7,2
7,7	8,7	8,2
8,7	9,7	9,2
9,7	10,7	10,2

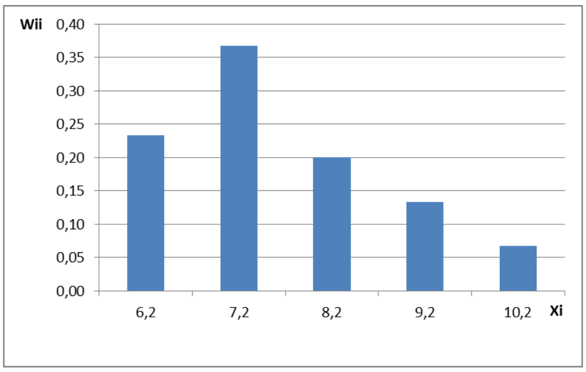
Далее подсчитываем частоты по каждому интервалу.

Правило: если варианта попадает на «стык» интервалов, то её следует относить в правый интервал. У нас такая варианта встретилась одна: x=8,7 – и её нужно причислить к интервалу (8,7-9,7).

В результате получаем интервальный вариационный ряд, и, кроме того, рассчитываем относительные частоты $w_i = n_i/n$ по каждому интервалу.

Гистограмма относительных частот – это фигура, состоящая из прямоугольников, ширина которых равна длинам *частичных интервалов*, а высота – соответствующим *относительным частотам*:

Интервалы		***xi***	***ni***	***wi***
5,7	6,7	***6,2***	***7***	***0,2333***
6,7	7,7	***7,2***	***11***	***0,3667***
7,7	8,7	***8,2***	***6***	***0,2***
8,7	9,7	***9,2***	***4***	***0,1333***
9,7	10,7	***10,2***	***2***	***0,0667***
		Σ=	30	1



Вместе с гистограммой нередко требуют построить полигон.

П*олигон относительных частот* – это ломаная, соединяющая соседние точки (x_i, w_i), где x_i – середины интервалов.

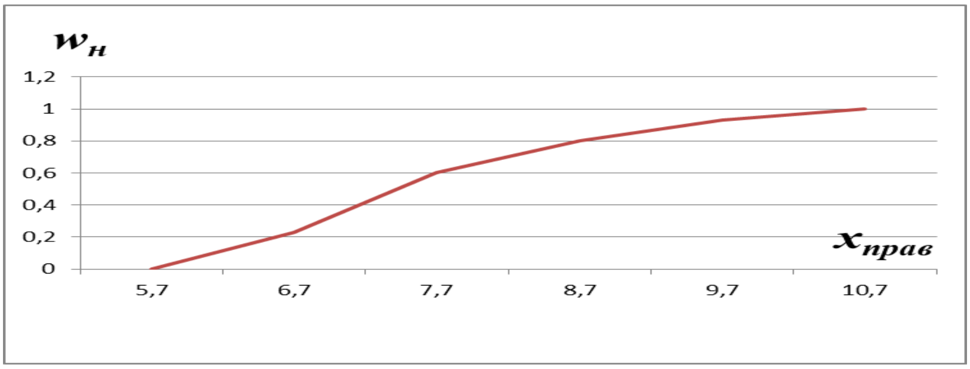
Эмпирическая функция распределения. Она определяется точно так же, как в дискретном случае: $F^*(x) = n_\Sigma/n$,

где n_Σ – количество вариант СТРОГО МЕНЬШИХ, чем «икс», который «пробегает» все значения от «минус» до «плюс» бесконечности.

Но вот построить её для интервального ряда намного проще. Находим накопленные относительные частоты:

x_i	n_i	w_i	$w_н$
6,2	7	0,23	0,23
7,2	11	0,37	0,60
8,2	6	0,20	0,80
9,2	4	0,13	0,93
10,2	2	0,07	1,00
Σ=	30	1	

И строим кусочно-ломаную линию, с промежуточными точками $(x_{прав}, w_н)$, где $x_{прав}$ – правые концы интервалов, а $w_н$ – относительная частота, которая успела накопиться на всех «пройденных» интервалах:



При этом $F^*(x)=0$ если $x \le 5{,}7$ и $F^*(x)=1$ если если $x>10{,}7$.

Данная функция ***не убывает***, принимает значения из промежутка $0 \le F^*(x) \le 1$ и, кроме того, для ИВР она ещё и **непрерывна**.

Эмпирическая функция распределения является аналогом **функции распределения НСВ** и *приближает* теоретическую функцию $F^*(x)$, которую теоретически, а иногда и практически можно построить по всей генеральной совокупности.

Варианты задания 1

№ вар	*Цена ед. товара в разл. торг. точках, ден.ед.*														
1	**9,58**	**8,33**	**6,99**	**8,25**	**9,7**	**6,61**	**5,73**	**6,68**	**8,77**	**9,75**	**8,51**	**5,73**	**9,67**	**5,41**	**8,83**
2	**9,22**	**6,76**	**8,34**	**9,93**	**8,37**	**5,07**	**9,68**	**6,18**	**5,99**	**8,13**	**8,38**	**5,75**	**7,71**	**9,03**	**8,04**
3	**8,23**	**9,3**	**7,17**	**7,58**	**6,97**	**9,14**	**8,49**	**8,36**	**8,5**	**7,67**	**6,47**	**9,2**	**8,89**	**9,46**	**8,79**
4	**8,91**	**9,64**	**7,62**	**7,12**	**5,84**	**6,37**	**7,52**	**5,47**	**9,92**	**9,88**	**8,77**	**5,61**	**9,04**	**7,09**	**5,44**
5	**6,97**	**5,87**	**7,6**	**9,43**	**8**	**7,37**	**9,91**	**5,32**	**6,97**	**6,88**	**9,13**	**6,88**	**9,26**	**5,47**	**5,93**
6	**8,25**	**9,6**	**9,52**	**7,22**	**5,16**	**5,67**	**8,37**	**6,02**	**8,62**	**7,61**	**6,36**	**9,59**	**7,07**	**7,98**	**7,45**
7	**5,92**	**8,16**	**5,24**	**9,03**	**8,22**	**9,28**	**7,81**	**8,67**	**6,7**	**6,16**	**5,81**	**6,24**	**6,23**	**6,39**	**9,62**
8	**7,72**	**5,69**	**5,21**	**7,04**	**5,84**	**6,43**	**5,05**	**6,26**	**5,55**	**8,83**	**7,01**	**9,29**	**8,14**	**5,17**	**9,5**
9	**5,06**	**5,22**	**8,36**	**8,89**	**9,85**	**8,01**	**9,59**	**6,48**	**7,62**	**7,46**	**7,98**	**7,95**	**7,84**	**5,51**	**7,89**
10	**5,03**	**5,65**	**5,58**	**5,87**	**9,46**	**6,59**	**5,93**	**5,63**	**5,08**	**8,95**	**7,36**	**6,84**	**5,15**	**6,36**	**7,48**
11	**9,14**	**5,43**	**5,6**	**6,41**	**9,56**	**7,83**	**5,91**	**6,67**	**6,25**	**6,26**	**9,58**	**5,99**	**9,6**	**5,45**	**6,92**
12	**7,01**	**8,64**	**9,03**	**8,76**	**7,37**	**5,36**	**7,21**	**6,39**	**5,7**	**6,35**	**5,15**	**6,31**	**8,86**	**5,88**	**5,74**
13	**7,88**	**8,98**	**6,6**	**5,66**	**9,06**	**9,29**	**7,33**	**6,18**	**6,47**	**5,05**	**5,33**	**5,64**	**7,52**	**5,98**	**5,73**
14	**8,07**	**7,95**	**5,91**	**7,07**	**5,67**	**7,68**	**6,34**	**8,98**	**8,54**	**5,64**	**9,58**	**5,29**	**8,04**	**6,94**	**7,1**
15	**7,03**	**5,32**	**5,28**	**9,07**	**8,03**	**5,28**	**6,46**	**5,95**	**8,25**	**7,34**	**5,28**	**9,02**	**9,78**	**7,96**	**6,83**
16	**9,1**	**9,64**	**9,66**	**6,62**	**5,94**	**8,24**	**7,03**	**6,72**	**9,79**	**7,83**	**5,84**	**8,79**	**7,15**	**5,4**	**5,42**
17	**8,37**	**7,66**	**8,77**	**8,34**	**7,14**	**9,73**	**5,93**	**9,37**	**7,68**	**6,94**	**7,73**	**7,76**	**7,55**	**7,57**	**5,89**
18	**8,03**	**9,87**	**7,38**	**9,29**	**9,42**	**8,65**	**6,84**	**6,03**	**7,39**	**5,4**	**5,67**	**9,51**	**6,4**	**6,38**	**7,03**
19	**9,51**	**9,73**	**6,24**	**5,89**	**6,92**	**9,67**	**7,95**	**8,93**	**6,58**	**5,83**	**7,16**	**9,28**	**9,44**	**7,98**	**5,23**
20	**7,21**	**9,98**	**9,54**	**7,37**	**7,01**	**9,25**	**9,24**	**5,97**	**9,39**	**5,17**	**9,87**	**7,71**	**6,56**	**7,04**	**8,26**
21	**8,21**	**9,29**	**8,99**	**5,86**	**9,36**	**5,99**	**6,9**	**8,75**	**9,42**	**6,2**	**6,67**	**5,53**	**6,02**	**6,29**	**5,39**
22	**8,62**	**9,27**	**8,28**	**9,08**	**9,55**	**6,69**	**7,38**	**5,8**	**7,18**	**6,44**	**9,29**	**9,74**	**6,09**	**8,78**	**9,52**
23	**7,62**	**5,46**	**6,42**	**9,83**	**7,67**	**7,19**	**8,93**	**5,3**	**7,84**	**5,38**	**7,98**	**8**	**6,81**	**8,61**	**7,26**
24	**5,49**	**9,83**	**9,63**	**6,48**	**5,62**	**8,95**	**9,74**	**6,51**	**8,66**	**6,01**	**8,75**	**7,2**	**9,27**	**7,63**	**5,98**
25	**7,59**	**6,87**	**9,47**	**7,77**	**9,37**	**5,14**	**6,7**	**6,07**	**5,36**	**8,63**	**9,82**	**7,23**	**7,79**	**5,74**	**6,17**

Задание 2

Выборочная проверка партии чая, поступившего в торговую сеть, дала следующие результаты:

Вес, грамм, x_i	48-49	49-50	50-51	51-52
Количество пачек, n_i	20	50	20	10

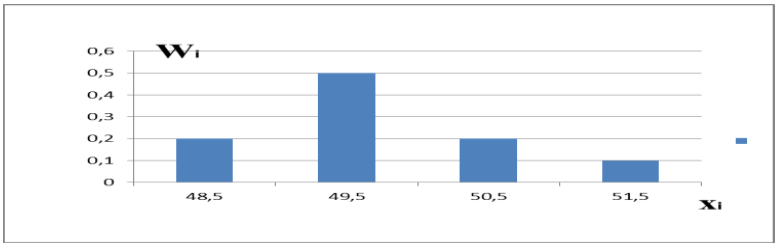
Требуется построить гистограмму и полигон относительных частот, эмпирическую функцию распределения.

Методика выполнения

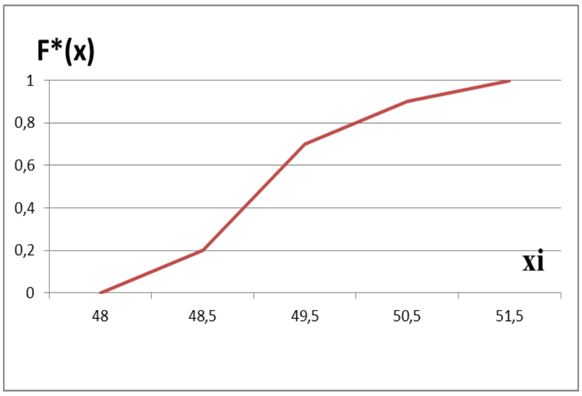
Заполним расчётную таблицу.

Интервалы		n_i	x_i	w_i	$w_н$
48	49	20	48,5	0,2	0,2
49	50	50	49,5	0,5	0,7
50	51	20	50,5	0,2	0,9
51	52	10	51,5	0,1	1
Суммы:		100		1	

Построим гистограмму относительных частот:



Построим эмпирическую функцию распределения:



Варианты задания 2

№ вар.	Вес, грамм, X_i				
	47-48	48-49	49-50	51-52	52-53
	Количество пачек, ni				
1	22	17	41	32	21
2	44	37	12	38	36
3	19	13	20	22	35
4	32	17	44	41	29
5	16	42	28	27	34
6	42	15	47	13	21
7	43	47	37	12	13
8	15	30	35	50	26
9	48	45	35	41	14
10	45	41	41	45	25
11	37	25	24	42	20
12	41	23	17	44	37
13	37	38	39	10	13
14	45	18	50	31	32
15	42	42	50	29	30
16	31	22	22	12	10
17	46	33	49	42	16
18	31	38	23	37	28
19	50	30	46	12	43
20	19	24	11	31	16
21	31	38	19	28	16
22	43	16	41	23	47
23	38	50	37	50	48
24	37	35	47	28	10
25	27	49	46	46	49