ПРАКТИЧЕСКОЕ ЗАНЯТИЕ №9

Дискретные и интервальные вариационные ряды и их числовые характеристики

На занятии рассматриваются вопросы:

1. Генеральная и выборочная совокупности.
2. Понятие вариационного ряда и его графическое изображение.
3. Числовые характеристики вариационных рядов: размах вариации, средняя арифметическая (выборочная средняя), выборочная дисперсия, выборочное среднее квадратическое отклонение, коэффициент вариации, мода, медиана.
4. Понятие оценки параметров. Свойства оценок.
5. Точечные оценки параметров генеральной совокупности.
6. Понятие интервального оценивания параметров генеральной совокупности.

Ответьте на вопросы для самоконтроля знания основ теоретического материала и решите предложенные задачи:

1. Что называют генеральной совокупностью?

Множество всех значений некоторой величины, то есть совокупность *всех* объектов, по которым проводится статистическое исследование, называется ***генеральной совокупностью***.

2. Что называют объемом генеральной совокупности?

Количество объектов генеральной совокупности N называется ***объемом генеральной совокупности.***

3. Что называют выборочной совокупностью (выборкой)?

Часть объектов генеральной совокупности, используемая для исследования, называется ***выборочной совокупностью*** (***выборкой***).

4. Что называют объемом выборки?

Число n объектов этой выборочной совокупности называется ***объемом выборки.***

Пусть исследуемый количественный признак X — **дискретная СВ.**

5. Что называют вариантами?

Наблюдаемые значения x_i признака X называются ***вариантами***.

6. Что такое ранжирование выборки?

Ранжирование выборки – это упорядочивание чисел x_i по возрастанию.

7. Что называют дискретным вариационным рядом?

Последовательность вариант, записанных в возрастающем порядке, называется ***дискретным вариационным рядом.***

Пример 1.

Пусть на 1 курсе архитектурно-строительного факультета обучается 200 человек. Случайным образом для исследования взяты результаты по контрольной работе (в баллах по 5-ти бальной шкале) 25 студентов. Оказалось, что 2 балла – у 5 студентов, 3 балла – у 10 студентов, 4 балла – у 7 студентов, 5 баллов – у 3 студентов.

Задание 1.

Для примера 1 определите:

Чему равен объем генеральной совокупности?

Чему равен объем выборки?

Составьте вариационный ряд признака X – балла за контрольную работу.

Решение:

Объем генеральной совокупности $N = 200$.

Объем выборки $n = 5 + 10 + 7 + 3 = 25$.

Вариантами признака X будут: $x_1 = 2$, $x_2 = 3$, $x_3 = 4$, $x_4 = 5$.

2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5 — дискретный вариационный ряд.

8. Что называют частотой варианты x_i?

Частотой варианты x_i называется число n_i, показывающее, сколько раз эта варианта встречается в выборке.

9. Что называют частостью (относительной частотой) варианты x_i?

Частостью (относительной частотой) варианты x_i называется число $w_i = \dfrac{n_i}{n}$.

Задание 2.

Для примера 1 составьте таблицу частот и таблицу относительных частот.

Решение:

Таблица частот и относительных частот (таблица 1)

Таблица 1

Варианта x_i	2	3	4	5
Частота n_i	5	10	7	3
Относительная частота w_i	$\frac{5}{25}$	$\frac{10}{25}$	$\frac{7}{25}$	$\frac{3}{25}$

Проверка: $\sum_{i=1}^{k} n_i = 25$, $\sum_{i=1}^{k} w_i = 1$.

10. Что называют полигоном частот?

Полигон частот — ломаная, соединяющая точки плоскости с координатами $(x_i; n_i)$, где $i = 1;2;...k$.

11. Что называют полигоном относительных частот?

Полигон относительных частот — ломаная, соединяющая точки плоскости с координатами $(x_i; w_i)$, где $i = 1;2;...k$.

Задание 3.

Для примера 1 постройте ***самостоятельно*** полигон частот и полигон относительных частот.

Решение:

12. Что называют накопленной (кумулятивной) частотой?

Накопленная (кумулятивная) частота m_i — количество вариант, значения которых меньше n_i.

$$m_i = n_1 + n_2 + \ldots + n_{i-1}.$$

13. Что называют относительной накопленной (кумулятивной) частотой?

Относительная накопленная частота v_i — отношение накопленной частоты к общему объему выборки, то есть

$$v_i = \frac{m_i}{n}.$$

Задание 4.

Для примера 1 составьте таблицу накопленных частот и таблицу относительных накопленных частот.

Решение:

Используем таблицу частот (часть таблицы 1):

Таблица 1

Варианта x_i	2	3	4	5
Частота n_i	5	10	7	3

Составим таблицу накопленных частот и относительных накопленных частот (таблица 2)

Таблица 2

Варианта x_i	2	3	4	5	6
Накопленная частота m_i	0	5	15	22	25
Относительная накопленная частота $v_i = \frac{m_i}{n}$	0	$\frac{5}{25} = 0{,}2$	$\frac{15}{25} = 0{,}6$	$\frac{22}{25} = 0{,}88$	$\frac{25}{25} = 1$

14. Что называют кумулятой частот?

Кумулята частот — ломаная, соединяющая точки с координатами $(x_i; m_i)$, где $i = 1;2;...k$.

15. Что называют кумулятой относительных частот?

Кумулята относительных частот — ломаная, соединяющая точки с координатами $(x_i; v_i)$, где $i = 1;2;...k$.

Задание 5.

Для примера 1 постройте ***самостоятельно*** кумуляту частот и кумуляту относительных частот.

Решение:

16. Что называют эмпирической функцией распределения?

Эмпирической функцией распределения $F^*(x)$ называют функцию, значение которой в точке x равно накопленной относительной частоте, то есть

$$F^*(x) = v_x = \frac{m_x}{n}.$$

Задание 5.

Для примера 1 составьте эмпирическую функцию распределения и постройте ее график.

Решение:

Используем таблицу относительных накопленных частот (таблица 2)

Таблица 2

Варианта x_i	2	3	4	5	6
Относительная накопленная частота $v_i = \frac{m_i}{n}$	0	$\frac{5}{25} = 0{,}2$	$\frac{15}{25} = 0{,}6$	$\frac{22}{25} = 0{,}88$	$\frac{25}{25} = 1$

Найдем эмпирическую функцию распределения:

$$F^*(x) = \begin{cases} 0, \text{ если } x \in (-\infty; 2], \\ 0{,}2, \text{ если } x \in (2; 3], \\ 0{,}6, \text{ если } x \in (3; 4], \\ 0{,}88, \text{ если } x \in (4; 5], \\ 1, \text{ если } x \in (5; +\infty). \end{cases}$$

График эмпирической функции распределения постройте самостоятельно.

Пусть количественный признак X — **непрерывная СВ**, принимающая значения из некоторого промежутка.

17. Как составляется интервальный вариационный ряд для количественного признака *X*, являющегося непрерывной случайной величиной, или при большом числе наблюдений и большом числе вариант?

При исследовании непрерывной СВ или при большом числе наблюдений и большом числе вариант диапазон наблюдаемых данных делят на частичные интервалы $[x_{i-1}; x_i)$. Если варианта находится на границе интервала, то ее присоединяют к правому интервалу.

Рекомендуется оптимальную длину частичного интервала (интервальную разность) определять по формуле Стерджерса:

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \lg n}$$

Тогда количество интервалов:

$$k = \frac{x_{\max} - x_{\min}}{h}.$$

Получим *равновеликий* интервальный ряд. В интервальном вариационном ряде могут встречаться интервалы разной длины. Такие ряды называют *неравновеликими*.

При малом объеме выборки (20–40 вариант) намечают 5–6 интервалов. При этом левая граница первого интервала и правая граница последнего интервала может не совпадать соответственно с $x_{\min}$ и с $x_{\max}$, но они должны быть как можно ближе к ним и включать их в себя.

Последовательность частичных интервалов, записанных в возрастающем порядке, называется ***интервальным вариационным рядом.***

Первый и последний интервалы могут не иметь соответственно левой и правой границы, то есть быть открытыми. В процессе обработки данных открытые интервалы приходится условно закрывать. При этом величину первого интервала чаще всего принимают равной величине второго, а величину последнего — величине предпоследнего.

Пример 2.

Из крупной партии растений произведена случайная выборка, получено 20 вариант длины стебля (в см): 35,9; 35,3; 42,7; 46,2; 25,9; 35,3; 33,4; 27,0; 35,9; 38,8; 33,7; 38,6; 40,9; 35,5; 44,1; 37,4; 34,2; 30,8; 38,4; 31,3.

Задание 6.

Для примера 2 постройте интервальный вариационный ряд.

Решение

Запишем исходные данные в виде ранжированного ряда, т.е. располагая их в порядке возрастания:

25,9; 27,0; 30,8; 31,3; 33,4; 33,7; 34,2; 35,3; 35,3; 35,5; 35,9; 35,9; 37,4; 38,4; 38,6; 38,8; 40,9; 42,7; 44,1; 46,2.

Максимальное значение признака составляет $x_{\max} = 46{,}2$ см, а минимальное – $x_{\min} = 25{,}9$ см. Разница между ними равна 20,3 см.

Так как объем выборки небольшой разобьем интервал наблюдаемых значений на 5 частичных интервалов.

Возьмем длину каждого частичного интервала $h_i = 5$ см.

Получаем интервальный вариационный ряд:
$[25;30)$, $[30;35)$, $[35;40)$, $[40;45)$, $[45;50]$.

18. Что называют частотой *i*-го интервала?

Частотой i-го интервала n_i называют число наблюдаемых значений в этом интервале.

19. Что называют частостью или относительной частотой *i*-го интервала?

Частостью или относительной частотой i-го интервала w_i называют отношение

$$w_i = \frac{n_i}{n}.$$

Задание 7.

Для примера 2 составьте таблицу частот и таблицу относительных частот.

Решение:

$w_i = \frac{n_i}{n}$, $n = 20$ – объем выборки.

Таблица частот и относительных частот (таблица 3):

Таблица 3

Интервал значений длины стебля	$[25;30)$	$[30;35)$	$[35;40)$	$[40;45)$	$[45;50]$	
Частота n_i	2	5	9	3	1	$\sum n_i = 20$
Относительная частота $w_i = \frac{n_i}{n}$	0,10	0,25	0,45	0,15	0,05	$\sum w_i = 1{,}00$

20. Что называют гистограммой частот? Чему равна площадь гистограммы частот?

Гистограмма частот — ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h_i, а высоты равны ***плотности частот*** $\frac{n_i}{h_i}$.

Площадь гистограммы частот $S_ч$ равна объему выборки:

$$S_ч = n.$$

Задание 8.

Для примера 2 постройте гистограмму частот.

Решение:

Для построения гистограммы частот составим таблицу плотности частот (таблица 4), используя таблицу частот (таблица 3):

Таблица 4

Интервал значений длины стебля	$[25;30)$	$[30;35)$	$[35;40)$	$[40;45)$	$[45;50]$
Частота n_i	2	5	9	3	1
Плотность частоты $\frac{n_i}{h_i}$	$\frac{2}{5}$	1	$\frac{9}{5}$	$\frac{3}{5}$	$\frac{1}{5}$

Учитываем, что $h_i = 5$.

Гистограмму частот постройте ***самостоятельно****.*

21. Что называют гистограммой относительных частот? Чему равна площадь гистограммы относительных частот?

Гистограмма относительных частот — ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h_i, а высоты равны ***плотности относительных частот*** $\frac{w_i}{h_i}$.

Площадь гистограммы относительных частот $S_{отн.ч}$ равна единице:

$$S_{отн.ч} = 1.$$

Задание 9.

Для примера 2 постройте гистограмму относительных частот.

Решение:

Для построения гистограммы относительных частот составим таблицу плотности относительных частот (таблица 5), используя таблицу относительных частот (таблица 3):

Таблица 5

Интервал значений длины стебля	$[25;30)$	$[30;35)$	$[35;40)$	$[40;45)$	$[45;50]$
Относительная частота $w_i = \frac{n_i}{n}$	0,10	0,25	0,45	0,15	0,05
Плотность относительной частоты $\frac{w_i}{h_i}$	0,02	0,05	0,09	0,03	0,01

Учитываем, что $h_i = 5$.

Гистограмму относительных частот постройте ***самостоятельно***.

22. Как вычисляется накопленная (кумулятивная) частота m_i для интервального ряда?

Накопленная (кумулятивная) частота m_i для интервального ряда вычисляется следующим образом:

Накопленная частота нижней границы первого интервала равна нулю.

Накопленная частота верхней границы первого интервала равна частоте этого интервала. Накопленная частота верхней границы второго интервала равна сумме частот первого и второго интервалов. Накопленная частота верхней границы третьего интервала равна сумме частот первого, второго и третьего интервалов. Далее аналогично.

23. По какой формуле находятся относительные накопленные частоты v_i интервального ряда?

Относительные накопленные частоты v_i интервального ряда находятся по формуле $v_i = \dfrac{m_i}{n}$.

Задание 10.

Для примера 2 составьте таблицу накопленных (кумулятивных) частот и относительных накопленных (кумулятивных) частот.

Решение:

Составим таблицу накопленных (кумулятивных) частот и относительных накопленных (кумулятивных) частот (таблица 6), используя таблицу частот и относительных частот (таблица 3):

Таблица 3

Интервал значений длины стебля	$[25; 30)$	$[30; 35)$	$[35; 40)$	$[40; 45)$	$[45; 50]$	
Частота n_i	2	5	9	3	1	$\sum n_i = 20$
Относительная частота $w_i = \dfrac{n_i}{n}$	0,10	0,25	0,45	0,15	0,05	$\sum w_i = 1{,}00$

Таблица 6

x_i	25	30	35	40	45	50
Накопленная частота m_i	0	2	2+5=7	7+9=16	16+3=19	19+1=20
Относительная накопленная частота $v_i = \dfrac{m_i}{n}$	0	0,1	0,35	0,8	0,95	1

24. Какие вы знаете числовые характеристики вариационного ряда и как они вычисляются?

Числовые характеристики вариационного ряда:

1. Размахом вариации называется число

$$R = x_{\max} - x_{\min}.$$

При этом для интервального вариационного ряда $x_{\max}$ — правый конец последнего частичного интервала, а $x_{\min}$ — левый конец первого частичного интервала.

2. Средняя арифметическая (выборочная средняя).

$$\overline{x}_в = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i n_i ,$$

где x_i — i-ая варианта признака для дискретного вариационного ряда или середина i-го интервала для интервального вариационного ряда,
n_i — частота i-го значения признака,
k — число значений признака;
или

$$\overline{x}_в = \sum_{i=1}^{k} x_i w_i ,$$

где w_i — относительная частота i-го значения признака.

Выборочная средняя характеризует среднее значение признака X генеральной совокупности.

3. Выборочная дисперсия.

$$D_в = \frac{1}{n} \cdot \sum_{i=1}^{k} \left(x_i - \overline{x}_в\right)^2 n_i$$

или

$$D_в = \sum_{i=1}^{k} \left(x_i - \overline{x}_в\right)^2 w_i .$$

Также применяют формулы:

$$D_в = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i^2 n_i - \left(\overline{x}_в\right)^2$$

или

$$D_в = \sum_{i=1}^{k} x_i^2 w_i - \left(\overline{x}_в\right)^2 .$$

Дисперсия характеризует отклонение от средней в квадратных единицах измерения признака.

4. Исправленная выборочная дисперсия.

$$S^2 = \frac{n}{n-1} D_в$$

5. Выборочное среднее квадратическое отклонение.

$$\sigma_в = \sqrt{D_в}\ .$$

6. Исправленное выборочное среднее квадратическое отклонение.

$$S = \sqrt{S^2}\ .$$

5. Коэффициент вариации.

$$V = \frac{\sigma_в}{\left|\overline{x_в}\right|} \cdot 100\% \quad (\overline{x_в} \neq 0).$$

$$V = \frac{S}{\left|\overline{x_в}\right|} \cdot 100\%$$

6. Модой вариационного ряда *Mo* называется то из значений x_1, x_2, $x_3, \ldots,\ x_n$, которому соответствует наибольшая частота.

7. ***Медиана*** *Me* — это значение варианты, которая является серединой вариационного ряда.

Если количество вариант четное, то медиана вычисляется как среднее двух вариант, находящихся в середине множества.

25. Как оценить качество выборочной средней?

Для оценки качества выборочной средней используют среднее квадратическое отклонение выборочной средней $\sigma_{\bar{x}_в} = \frac{S}{\sqrt{n}}$ при $n \le 30$ или $\sigma_{\bar{x}_в} = \frac{\sigma}{\sqrt{n}}$ при $n > 30$.

Задание 9.

Для примера 2 вычислите выборочную среднюю $\overline{x}_в$, выборочную дисперсию $D_в$, исправленную выборочную дисперсию S^2, исправленное выборочное среднее квадратическое отклонение S, среднее квадратическое отклонение выборочной средней $\sigma_{\overline{x}_в}$, коэффициент вариации V.

Решение:

Таблица 7. Расчетная таблица выборочных характеристик

Варианта x_i	Частота варианты n_i	$x_i n_i$	$\left(x_i - \overline{x}_в\right)^2 n_i$
27,5	2	55	162
32,5	5	162,5	80
37,5	9	337,5	9
42,5	3	127,5	108
47,5	1	47,5	121
$\sum$	20	730	480

Найдем выборочную среднюю:

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i ,$$

$$\overline{x}_в \approx \frac{1}{20}\cdot 730 \approx 36{,}50 \text{ см.}$$

Вычислим выборочную дисперсию:

$$D_в = \frac{1}{n}\cdot\sum_{i=1}^{k}\left(x_i - \overline{x}_в\right)^2 n_i ,$$

$$D_в \approx \frac{1}{20}\cdot 480 \approx 24{,}00 \text{ см}^2 .$$

Вычислим исправленную выборочную дисперсию:

$$S^2 = \frac{n}{n-1} D_в ,$$

$$S^2 \approx \frac{20}{20-1}\cdot 24 \approx 25{,}26 \text{ см}^2 .$$

Найдем исправленное выборочное среднее квадратическое отклонение:

$$S = \sqrt{S^2} ,$$

$$S \approx \sqrt{25{,}26} \approx 5{,}03 \text{ см.}$$

Вычислим среднее квадратическое отклонение выборочной средней:

$$\sigma_{\overline{x}_в} = \frac{S}{\sqrt{n}} ,$$

$$\sigma_{\bar{x}_в} \approx \frac{5{,}03}{\sqrt{20}} \approx 1{,}13 \text{ см.}$$

Найдем коэффициент вариации:

$$V = \frac{S}{\left|\bar{x}_в\right|} \cdot 100\%,$$

$$V \approx \frac{5{,}03}{\left|36{,}50\right|} \cdot 100\% \approx 13{,}78\%.$$

Так как $10\% < V < 20\%$, то изменчивость длины стебля следует считать средней.

26. Как найти доверительный интервал для оценки генеральной средней при $n \le 30$?

Доверительный интервал для математического оценки генеральной средней имеет вид:

$$\left(\overline{x}_в - \Delta;\ \overline{x}_в + \Delta\right),$$

где $\Delta = t_\gamma \cdot \sigma_{\overline{x}_в} = t_\gamma \cdot \dfrac{S}{\sqrt{n}}$ – точность оценки.

Значение t_γ определяется по таблице значений $t_\gamma = t(\gamma; n)$ в зависимости от надежности γ и объема выборки *n*.

Задание 9.

Для примера 2 с надежностью 95% укажите доверительный интервал для оценки генеральной средней $\overline{x}_Г$.

Решение:

Доверительный интервал для оценки генеральной средней определяется следующим неравенством

$$\overline{x}_в - t_\gamma \sigma_{\overline{x}_в} < \overline{x}_Г < \overline{x}_в + t_\gamma \sigma_{\overline{x}_в},$$

где величина t_γ определяется с помощью таблицы значений $t_\gamma = t(\gamma; n)$ в зависимости от надежности γ и объема выборки *n*.

В нашем примере

$$t_\gamma = t(0{,}95;\, 20) = 2{,}093 \approx 2{,}09.$$

Вычислим радиус доверительного интервала:

$$t_\gamma \sigma_{\overline{x}_в} \approx 2{,}09 \cdot 1{,}13 \approx 2{,}36 \text{ см.}$$

Таким образом, с надежностью 95% можно утверждать, что во всей партии растений средняя длина стебля (генеральная средняя) заключена в пределах от $\overline{x}_в - t_\gamma \sigma_{\overline{x}_в} \approx 36{,}50 - 2{,}36 \approx 34{,}14$ см до $\overline{x}_в + t_\gamma \sigma_{\overline{x}_в} \approx 36{,}50 + 2{,}36 \approx 38{,}86$ см.

Итак, доверительный интервал для оценки генеральной средней с надежностью 95 % задается неравенством

$$34{,}14 \text{ см} < \overline{x}_Г < 38{,}86 \text{ см}.$$