

2.7 Тема. Структурные средние

Цель. Знакомство с методами вычисления основных биометрических показателей количественных признаков при малых и больших выборках.

Медианой называют середину класса, который делит вариационный ряд на две части: одна имеет значение признака меньше, чем медиана, другая – больше. Медиана обозначается символом Me .

На небольших выборках определить медиану довольно легко. Для этого совокупность наблюдений ранжируют по возрастающим значениям признака, и если число членов ряда нечетное, то центральная варианта и будет его медианой. При четном числе членов ряда медиана определяется по полусумме двух соседних вариантов, расположенных в центре ряда.

Пример. Имеется следующий ряд числовых значений признака: 12 14 16 18 20 22 24 26 28. Медианой этого ряда будет центральная варианта, т.е. $Me=20$; в обе стороны от нее отстоит по четыре варианты, а для ряда – 6 8 10 12 14 16 18 20 22 24 – медианой будет полусумма двух его центральных членов, т.е. $Me=(14+16)/2=15$.

Вычисление медианы в больших выборках при неравномерном распределении вариантов по классам проводят по формуле:

$$Me = X_{Me} + k \left(\frac{\frac{n}{2} - P_s}{P_{Me}} \right), \quad (20)$$

где X_{Me} – нижняя граница интервала, в котором находится медиана, или полусумма соседних классовых вариантов промежутка, в котором должна быть медиана, если последняя определяется для выборки, распределенной в безынтервальный вариационный ряд; n – общее число вариантов в группе; P_s – число накопленных частот, стоящее перед медианным классом; P_{Me} – частота класса, в котором находится медиана; k – величина классowego промежутка.

Определение медианы проводится путем накопления частот от минимальной величины до величины, не превышающей полусуммы всех вариантов вариационного ряда. По этой величине устанавливается класс, в котором находится медиана.

Пример. Вычислим медиану ряда распределения кальция (мг%) в сыворотке крови павианов гамадрилов (таблица 2.7.1).

Таблица 2.7.1

Классы по содержанию кальция в сыворотке крови	Срединные значения классов	Частоты (p_i)	Накопленные частоты (p_s)
8,6-9,3	9,0	2	2
9,4-10,1	9,8	6	8

10,2-10,9	10,6	15	23
11,0-11,7	11,4	23	46
11,8-12,5	12,2	25	71
12,6-13,3	13,0	17	
13,4-14,1	13,8	7	
14,2-14,9	14,6	5	
Сумма			100

Величина $n/2=100/2=50$ находится между $p_s=46$ и $p_s=71$. По $p_s=71$ находим границы интервала, в котором находится медиана: от 11,8 до 12,5 с частотой $p_{Me}=25$; $k=0,8$. Подставляя известные величины в формулу (20), находим $Me=11,8+0,8(50-46)/25=11,8+0,128=11,93$. Такой же результат получается, если вместо нижней границы интервала, содержащего медиану, берется полусумма соответствующих классовых вариантов: $Me=(11,4+12,2)/2+0,8x(50-46)/25=11,8+0,128=11,93$.

Мода (M_o) – величина, которая встречается в данной совокупности наиболее часто. Класс с наибольшей частотой называется модальным. Для определения моды используется следующая формула:

$$M_o = x_h + k \left(\frac{p_2 - p_1}{2p_2 - p_1 - p_3} \right), \quad (21)$$

где x_h – нижняя граница модального класса, т.е. класса с наибольшей частотой (p_2); p_1 – частота класса, предшествующего модальному; p_3 – частота класса, следующего за модальным; k – ширина классического промежутка.

Пример. Определите моду ряда распределения кальция ($mg\%$) в сыворотке крови павианов гамадрилов. Из данных таблицы 2.7.1 следует, что нижняя граница класса с наибольшей частотой ($p_2=25$) равна $x_h=11,8$; частота предшествующего класса $p_1=23$, а следующего за модальным классом $p_3=17$; $k=0,8$. Подставляя эти величины в формулу (21), находим

$$M_o = 11,8 + 0,8 \left(\frac{25 - 23}{2 \times 25 - 23 - 17} \right) = 11,8 + 0,16 = 11,96.$$

Наряду с медианой и модой к структурным характеристикам вариационного ряда относятся так называемые квантили, отсекающие в пределах ряда определенную часть его членов (вариант). К ним относятся квартили, децили и перцентили (процентили).

Квартиль – величина, отсекающая $1/4$ членов ряда. Три квартиля – d_1 , d_2 и d_3 – делят весь вариационный ряд на четыре равночисленные части (кварти).

Дециль – величина, отделяющая $1/10$ всех членов ряда. Девять децилей делят вариационный ряд на десять равных частей. Если же речь идет о сотых долях вариант в общем ряду распределения, такая величина, обозначаемая символом P_i , называется перцентилем или процентилем. 99

перцентилей делят всю совокупность наблюдений на 100 равночисленных частей.

В практике используются обычно перцентили: $P_3, P_{10}, P_{25}, P_{50}, P_{75}, P_{90}, P_{97}$, причем 50-й перцентиль равен медиане, т.е. $P_{50} = Me$; он также соответствует второму квартилю (d_2) и пятому децилю данного распределения. Между P_{25} и P_{75} , и соответственно между d_1 и d_3 , находится 50% всех членов совокупности.

Любой перцентиль определяется рядом последовательных действий, которые можно выразить в виде следующей формулы:

$$P_i = X_h + k \left(\frac{K - p_s}{p} \right), \quad (22)$$

где X_h = нижняя граница класса, содержащего перцентиль P_i ; она определяется по величине $K = P_i n / 100$, превосходящей (или равной) числу p_s в ряду накопленных частот. Здесь P_i – выбранный перцентиль; p – частота класса, содержащего перцентиль P_i ; k – величина классового интервала; n – общее число наблюдений, или объем выборки. Таким образом, как и при вычислении медианы, определение перцентилей опирается на кумуляцию частот вариационного ряда, которая производится в направлении от минимальных к максимальным значениям признака.

Задание 1. Измерения 94 диаметров сосны (см) распределились следующим образом:

X.....	16	20	24	28	32	36	40
P.....	4	7	8	28	20	18	9
P_i	4	11	19	47	67	85	94

Определите медиану этого распределения.

Задание 2. Найдите медиану ряда распределения численности поросят в 64 пометах свиноматок.

Количество поросят в пометах (w)	Число случаев (p)	Накопленные частоты (p_i)	Количество поросят в пометах (w)	Число случаев (p)	Накопленные частоты (p_i)
5	4	4	10	9	
6	7	11	11	6	
7	13	24	12	3	
8	15	39			
9	7		Сумма	64	-

Задание 3. Найдите 50-й перцентиль для распределения численности поросят в 64 пометах свиноматок:

X.....	5	6	7	8	9	10	11	12
--------	---	---	---	---	---	----	----	----

P_i	4	7	13	15	7	9	6	3
P_s	4	11	24	39	46	55	61	64

Контрольные вопросы.

1. Дать характеристику структурным средним.
2. Чем отличается медиана от моды?
3. Как вычисляет перцентиль?