

Выборочное наблюдение

Практическое занятие

Крестьянские хозяйства подразделяются по процентам земельных угодий следующим образом:

Земельные угодия	Число хозяйств
До 5	40
6-10	50
11-20	70
21 и более	20

Найти:

1. Средний размер земельных угодий.
2. Показатели вариации: размах, среднее линейное, среднее квадратическое отклонение, коэффициент вариации.
Оценить количественную однородность совокупности.

Для расчета требуемых показателей следует перейти от вариационного ряда к дискретному. Для этого находится середина каждого интервала. Расчет показателей легче выполнять в таблице

Земельные угодия, га	Число хозяйств, ед.	Середина интервала, x_i	$x_i \cdot f_i$				Накопленные частоты
До 5	40	2,5	40*2,5				
6-10	50	8	50*8				
11-20	70	15,5	70*15,5				
21 и более	20	25	20*25				
Итого	180	-					

Для расчета требуемых показателей следует перейти от вариационного ряда к дискретному. Для этого находится середина каждого интервала. Расчет показателей легче выполнять в таблице

Земельные угодия, га	Число хозяйств, ед.	Середина интервала, x_i	$x_i \cdot f_i$				Накопленные частоты
До 5	40	2,5	100				
6-10	50	8	400				
11-20	70	15,5	1085				
21 и более	20	25	500				
Итого	180	-	2085				

Земельные угодия, га	Число хозяйств, ед.	Середина интервала, x_i	$x_i f_i$				Накопленные частоты
До 5	40	2,5	100				
6-10	50	8	400				
11-20	70	15,5	1085				
21 и более	20	25	500				
Итого	180	-	2085				

Средний размер земельных угодий определяется

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i},$$

где \bar{x} - среднее значение признака;

x - срединное значение интервала, в котором изменяется варианта (значение) осредняемого признака;

f_i - частота, с которой встречается данное значение осредняемого признака.

Земельные угодия, га	Число хозяйств, ед.	Середина интервала, x_i	$x_i \cdot f_i$				Накопленн ые частоты
До 5	40	2,5	100				
6-10	50	8	400				
11-20	70	15,5	1085				
21 и более	20	25	500				
Итого	180	-	2085				

Средний размер земельных угодий определяется

$$\bar{x} = \frac{2085}{180} = 11.6$$

Земельные угодия, га	Число хозяйств, ед.	Середина интервала, x_i	$x_i \cdot f_i$				Накоплен ные частоты
До 5	40	2,5	100	9,1			
6-10	50	8	400	3,6			
11-20	70	15,5	1085	3,9			
21 и более	20	25	500	13,4			
Итого	180	-	2085				

Средний размер земельных угодий определяется

$$\bar{x} = \frac{2085}{180} = 11.6$$

Земельные угодия, га	Число хозяйств, ед.	Середина интервала, x_i	$x_i \cdot f_i$				Накопленн ые частоты
До 5	40	2,5	100	9,1	9,1*40		
6-10	50	8	400	3,6			
11-20	70	15,5	1085	3,9			
21 и более	20	25	500	13,4			
Итого	180	-	2085				

Земельные угодия, га	Число хозяйств, ед.	Середина интервала, x_i	$x_i \cdot f_i$				Накоплен ные частоты
До 5	40	2,5	100	9,1	364		
6-10	50	8	400	3,6	180		
11-20	70	15,5	1085	3,9	273		
21 и более	20	25	500	13,4	268		
Итого	180	-	2085	-	1085		

Земельные угодия, га	Число хозяйств, ед.	Середина интервала, x_i	$x_i \cdot f_i$				Накопленн ые частоты
До 5	40	2,5	100	9,1	364	122,81	
6-10	50	8	400	3,6	180	648	
11-20	70	15,5	1085	3,9	273	1064,7	
21 и более	20	25	500	13,4	268	3591,2	
Итого	180	-	2085	-	1085	5426,71	

Земельные угодия, га	Число хозяйств, ед.	Середина интервала, x_i	$x_i \cdot f_i$				Накоплен ные частоты
До 5	40	2,5	100	9,1	364	122,81	40
6-10	50	8	400	3,6	180	648	90
11-20	70	15,5	1085	3,9	273	1064,7	160
21 и более	20	25	500	13,4	268	3591,2	180
Итого	180	-	2085	-	1085	5426,71	-

2. Рассчитаем указанные показатели вариации:

а) размах вариации:

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

б) среднее линейное отклонение:

$$\bar{l} = \frac{\sum |x_i - \bar{x}| \cdot f_i}{\sum f_i}$$

в) среднее квадратическое отклонение:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}}$$

г) коэффициент вариации:

$$V = \frac{\sigma}{\bar{x}} \cdot 100\%$$

- $R=21-5=16$ га
- Среднее линейное отклонение $1085/180=6,03$
- Среднее квадратическое отклонение $5,5$ га
- Коэффициент вариации $47,4 \%$

Вывод: крестьянские хозяйства количественно неоднородны по размеру земельных угодий, т.к. коэффициент вариации больше 33%

Рассчитаем структурные средние

- Найти моду и медиану

Мода

Мода — это наиболее часто встречающийся вариант ряда. Мода применяется, например, при определении размера одежды, обуви, пользующейся наибольшим спросом у покупателей. Модой для дискретного ряда является варианта, обладающая наибольшей частотой. При вычислении моды для интервального вариационного ряда необходимо сначала определить модальный интервал (по максимальной частоте), а затем — значение модальной величины признака по формуле:

$$M_0 = x_0 + n \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})},$$

где:

- M_0 — значение моды
- x_0 — нижняя граница модального интервала
- h — величина интервала
- f_m — частота модального интервала
- f_{m-1} — частота интервала, предшествующего модальному
- f_{m+1} — частота интервала, следующего за модальным

Медиана

Медиана — это значение признака, которое лежит в основе ранжированного ряда и делит этот ряд на две равные по численности части.

Для определения медианы **в дискретном ряду** при наличии частот сначала вычисляют полусумму частот $\frac{\sum f_i}{2}$, а затем определяют, какое значение варианта приходится на нее. (Если отсортированный ряд содержит нечетное число признаков, то номер медианы вычисляют по формуле:

$$M_e = (n(\text{число признаков в совокупности}) + 1)/2,$$

в случае четного числа признаков медиана будет равна средней из двух признаков находящихся в середине ряда).

При вычислении медианы **для интервального вариационного ряда** сначала определяют медианный интервал, в пределах которого находится медиана, а затем — значение медианы по формуле:

$$M_e = x_0 + h \frac{\frac{\sum f_i}{2} - S_{m-1}}{f_m},$$

где:

- M_e — искомая медиана
- x_0 — нижняя граница интервала, который содержит медиану
- h — величина интервала
- $\sum f_i$ — сумма частот или число членов ряда
- S_{m-1} - сумма накопленных частот интервалов, предшествующих медианному
- f_m — частота медианного интервала

Пример. Найти моду и медиану.

Возрастные группы	Число студентов	Сумма накопленных частот ΣS
До 20 лет	346	346
20 — 25	872	1218
25 — 30	1054	2272
30 — 35	781	3053
35 — 40	212	3265
40 — 45	121	3386
45 лет и более	76	3462
Итого	3462	

Решение:

В данном примере модальный интервал находится в пределах возрастной группы 25-30 лет, так как на этот интервал приходится наибольшая частота (1054).

Рассчитаем величину моды:

$$M_0 = x_0 + h \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} = 25 + 5 \frac{1054 - 872}{(1054 - 872) + (1054 - 781)}$$

= 27 лет.

Это значит что модальный возраст студентов равен 27 годам.

Вычислим медиану. Медианный интервал находится в возрастной группе 25-30 лет, так как в пределах этого интервала расположена варианта, которая делит совокупность на две равные части ($\Sigma f_i / 2 = 3462 / 2 = 1731$). Далее подставляем в формулу необходимые числовые данные и получаем значение медианы:

$$M_e = x_0 + h \frac{\frac{\Sigma f_i}{2} - S_{m-1}}{f_m} = 25 + 5 \frac{\frac{3462}{2} - 1218}{1054} = 27,4 \text{ года.}$$

Это значит что одна половина студентов имеет возраст до 27,4 года, а другая свыше 27,4 года.

Кроме моды и медианы могут быть использованы такие показатели, как квартили, делящие ранжированный ряд на 4 равные части, децили - 10 частей и перцентили — на 100 частей.