

Л.В. Шамшурина

Статистика

(Ряды распределения. Средние величины)

Ижевск 2007

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
ГОУВПО «УДМУРТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ИНСТИТУТ ЭКОНОМИКИ И УПРАВЛЕНИЯ

Статистика

(Ряды распределения. Средние величины)

Ижевск 2007

Методическая разработка обсуждена на заседании кафедры экономических методов управления.

«28» декабря 2006 г.

Зав. кафедрой ЭМУ

О.Д. Головина

Рекомендовано для использования в учебном процессе

«____» _____ 2007 г.

Председатель УМК

А.С. Баскин

Методическая разработка является частью учебно-методического комплекса по статистике. Она будет полезна тем, кто почувствовал потребность в помощи при изучении темы «Ряды распределения. Средние величины».

Шамшурина Л.В. Статистика: Ряды распределения. Средние величины. Для студентов экономических специальностей вузов. – Ижевск: Издательство ИЭиУ ГОУВПО «УдГУ», 2007, с. 20

© Л.В. Шамшурина, 2007
© ИЭиУ УдГУ, 2007

Содержание

1. Вариационные ряды	4
2. Средние величины	8
3. Показатели вариации	13
Задача 1	
Задача 2	
Задача 3	

1. Вариационные ряды

Первым шагом упорядочивания исходной информации является подсчет единиц изучаемой совокупности. В результате этого получаем ряд распределения.

Статистический ряд распределения – это упорядоченное распределение единиц изучаемой совокупности на группы по определенному варьирующему признаку. Он характеризует состав (структуру) изучаемого явления, позволяет судить об однородности совокупности, закономерности распределения и границах варьирования единиц совокупности (рис. 1).

Ряды распределения, построенные по атрибутивным признакам, называются *атрибутивными* (таблица 1).

Ряды распределения, построенные по количественному признаку (в порядке возрастания или убывания наблюдаемых значений), называются *вариационными* (таблицы 2-4).

Вариационный ряд (ВР) включает изучаемый признак и повторяемость признака (в виде частоты или частоты).

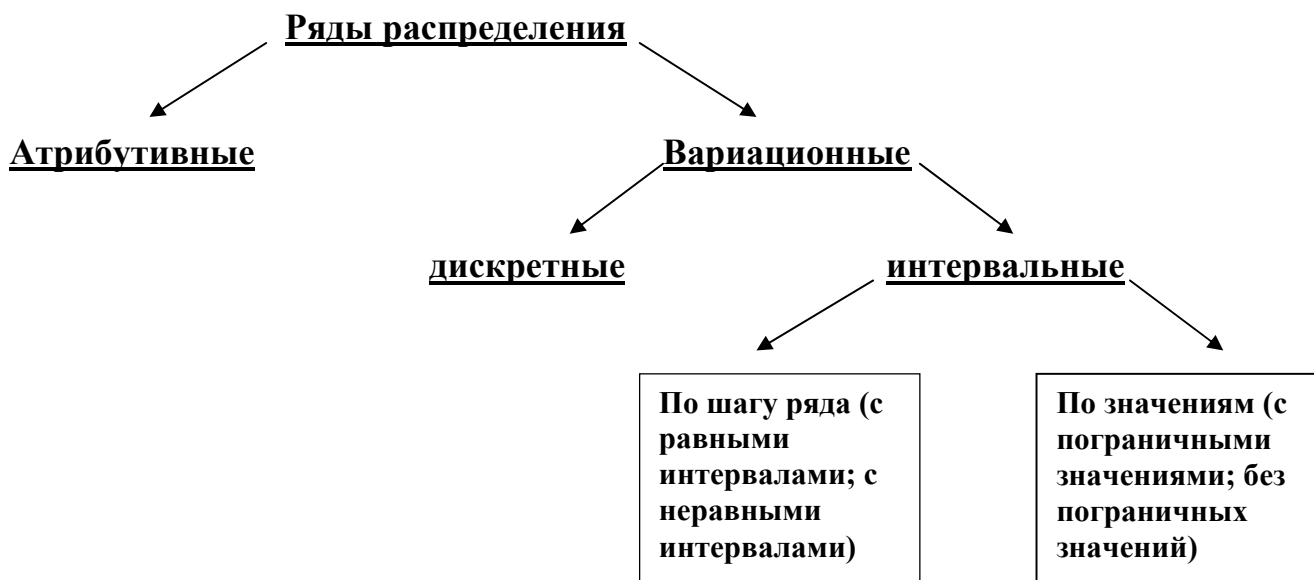


Рис. 1. Состав и классификация рядов распределения

Таблица 1

Численность незанятых трудовой деятельностью граждан, тыс.
чел.

Федеральный округ	2000 г.
1. Центральный	212,0
2. Северо-западный	141,1
3. Южный	144,0
4. Приволжский	247,3
5. Уральский	96,7
6. Сибирский	202,0
7. Дальневосточный	95,1

Таблица 2

Размер домохозяйства в РФ
(по микропереписи населения 1994 г., в процентах)

	Домохозяйства				
	Из 1 человека	Из 2 человек	Из 3 человек	Из 4 человек	Из 5 и более человек
Все домохозяйства:	19,2	26,2	22,6	20,5	11,5
городские	18,1	26,1	24,3	21,0	10,5
сельские	22,0	26,8	18,1	19,0	14,1

Таблица 3

Распределение населения УР по среднему денежному
доходу в 1999 г. (%)

	За год
Все население	100,0
в том числе со средним денежным доходом в месяц, руб.	
до 400,0	4,3
400,1 - 600,0	15,2
600,1 – 800,0	20,6
800,1 – 1000,0	18,7
1000,1 – 1200,0	14,1
1200,1 – 1600,0	16,0
1600,1 – 2000,0	6,6
Свыше 2000,0	4,5

Источник: Удмуртия в цифрах 1999г., с.28.

Таблица 4

Продолжительность ярмарок в России (по переписи 1894г.)

Продолжительность ярмарки, дни	Число ярмарок
Однодневная	10667
От 2 до 3 дней	4079
3 – 7	1347
7 – 12	267
12 – 20	184
20 – 30	63
30 и более	11

Источник: Россия: «Энциклопедический словарь, -Л.: Лениздат, 1991, с. 324.

Числовые значения количественного признака в вариационном ряду распределения называются *вариантами*.

Частоты – это численности отдельных вариантов или каждой группы вариационного ряда, то есть это числа, показывающие как часто встречаются те или иные варианты в ряду распределения. Сумма всех частот называется *объемом* совокупности и определяет число элементов всей совокупности.

Частоты – это частоты, выраженные в виде относительных величин (долях единиц или процентах). Сумма частот равна единице (или 100%). Замена частот частостями позволяет сопоставлять вариационные ряды с разным числом наблюдений.

В зависимости от характера вариации ВР могут быть дискретными (см. таблицу 2) и интервальными (см. таблицы 3,4).

Дискретные ВР основаны на дискретных (прерывных) признаках, имеющих только целые значения; на дискретных признаках, представленных в виде интервалов; интервальные – на непрерывных признаках.

Если необходимо построить ряд с заданным числом равных интервалов, то шаг в таком ряду считают по формуле:

$$h = \frac{X_{\max} - X_{\min}}{n}, \quad (1)$$

где h – шаг интервала;

$X_{\max(\min)}$ – соответственно наибольшее (наименьшее) значение признака;

n – число заданных равных интервалов.

Если количество интервалов не названо, то их можно рассчитать:

$$n = 1 + 3,322 \lg N, \quad (2)$$

где n – число интервалов в ряду;
 N – число единиц совокупности.

Центральное (срединное) значение интервала (X_{cp}) рассчитывается одним из методов:

$$X_{cp} = \frac{X_H + X_B}{2}, \quad (3)$$

или

$$X_{cp} = X_H + \frac{h}{2}, \quad (4)$$

или

$$X_{cp} = X_B - \frac{h}{2}. \quad (5)$$

2. Средние величины

1. Средняя арифметическая взвешенная

$$a) \bar{x} = \frac{\sum x \cdot f}{\sum f}, \text{ если } f_1 \neq f_2 \neq f_3 \neq \dots \neq const. \quad (6)$$

$$б) \bar{x} = \frac{\sum x \cdot d}{100}, \text{ если } d_1 \neq d_2 \neq d_3 \neq \dots \neq const. \quad (7)$$

$$в) \bar{x} = \sum x \cdot d., \text{ если } d_1 \neq d_2 \neq d_3 \neq \dots \neq const. \quad (8)$$

2. Средняя арифметическая простая

$$\text{а) } \bar{x} = \frac{\sum x}{n}, \text{ если } f_1 = f_2 = \dots = \text{const.}, \text{ то} \quad (9)$$

n – число уровней ряда.

$$\text{б) } \bar{x} = \frac{\sum x}{n}, \text{ если невозможно определить повторяемость, то} \quad (10)$$

n – размер изучаемой совокупности.

3. Средняя гармоническая простая

$$\bar{x} = \frac{n}{\sum \frac{1}{x}}, \quad (11)$$

n – число вариантов,

$\frac{1}{x}$ - отдельные варианты обратного признака, встречающиеся по одному разу.

4. Средняя гармоническая взвешенная

$$\bar{x} = \frac{\sum x \cdot f}{\sum \frac{x \cdot f}{x}}, \text{ когда известны: } x, \quad xf. \quad (12)$$

Структурные средние

1. Мода

Значение признака, которое чаще всего встречается (M_o):

$$M_o = X_n + h \frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)} \quad (13)$$

или

$$M_o = X_n + h \frac{d_2 - d_1}{(d_2 - d_1) + (d_2 - d_3)} \quad (14)$$

где X_n – нижняя граница модального интервала,

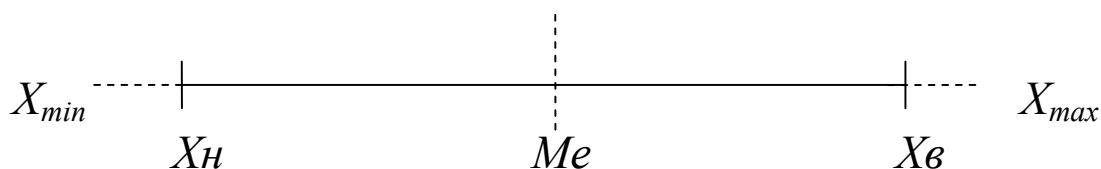
h – шаг модального интервала,

f_1, f_2, f_3 – соответственно частота предмодального, модального и послемодального интервала.

d_1, d_2, d_3 – соответственно частоты предмодального, модального и послемодального интервала.

2. Медиана

Значение признака находящегося в середине упорядоченного ряда (Me):



$$Me = X_H + h \frac{N_{Me} - S_{Me-1}}{f_{Me}} \quad (15)$$

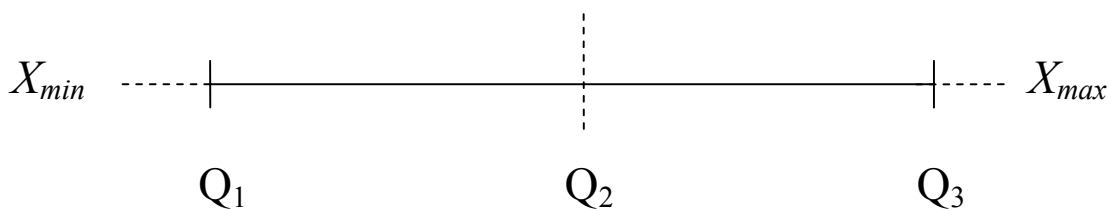
или

$$Me = X_H + h \frac{50\% - S_{Me-1}^d}{d_{Me}} \quad (16)$$

где X_H – нижняя граница медианного интервала,
 h – шаг медианного интервала ($X_B - X_H$),
 N_{Me} – порядковый номер медианы,
 S_{Me-1} – накопленная частота до медианного интервала,
 f_{Me} – частота медианного интервала,
 S_{Me-1}^d – накопленная частость до медианного интервала,
 d_{Me} – частость медианного интервала.

3. Квартили

Значения признака, делящие упорядоченный ряд на 4 равные части (Q):



Нижний квартиль:

$$Q_1 = X_{nQ_1} + h \frac{\frac{\sum f}{4} - S_{Q_1-1}}{f_{Q_1}} \quad (17)$$

Средний квартиль:

$$Q_2 = Me \quad (18)$$

Верхний квартиль:

$$Q_3 = X_{HQ_3} + h \frac{\frac{3 \cdot \Sigma f}{4} - S_{Q_3-1}}{f_{Q_3}}, \quad (19)$$

где X_{HQ_1} - нижняя граница интервала содержащая нижний квартиль,

h – шаг интервала ($X_v - X_n$),

S_{Q_1-1} - накопленная частота интервала, предшествующего интервалу, содержащему нижний квартиль,

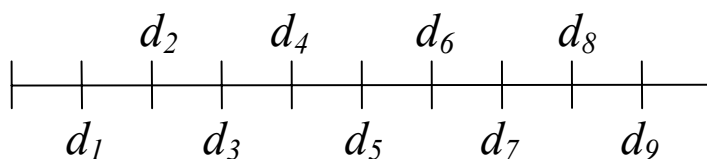
S_{Q_3-1} - накопленная частота интервала, предшествующего интервалу, содержащему верхний квартиль,

f_{Q_1} - частота интервала, содержащего нижний квартиль,

f_{Q_3} - частота интервала, содержащего верхний квартиль.

3. Децили

Варианты, делящие упорядоченный ряд на 10 равных частей (d).



Первый дециль:

$$d_1 = X_{Hd_1} + h \frac{\frac{\Sigma f}{10} - S_{d_1-1}}{fd_1} \quad (20)$$

Второй дециль:

$$d_2 = X_{Hd_2} + h \frac{\frac{2 \cdot \sum f}{10} - S_{d_2-1}}{fd_2} \quad (21)$$

...

Девятый дециль:

$$d_9 = X_{Hd_9} + h \frac{\frac{9 \cdot \sum f}{10} - S_{d_9-1}}{fd_9}, \quad (22)$$

где X_{Hd_1} - нижняя граница интервала, содержащего первый дециль,

X_{Hd_9} - нижняя граница интервала, содержащего девятый дециль,

h – шаг интервала,

S_{d_1-1} - накопленная частота интервала, предшествующего интервалу, содержащему первый дециль,

S_{d_9-1} - накопленная частота интервала, предшествующего интервалу, содержащему девятый дециль,

f_{d_1} - частота интервала, содержащего первый дециль,

f_{d_9} - частота интервала, содержащего девятый дециль.

3. Показатели вариации

1. Размах вариации: $R = X_{\max} - X_{\min}$ (23)

2. Среднее линейное отклонение:

а) для несгруппированных данных

$$\bar{d} = \frac{\sum |x - \bar{x}|}{n} \quad (24)$$

б) для сгруппированных данных

$$\bar{d} = \frac{\sum |x - \bar{x}| f}{\sum f} \quad (25)$$

3. Дисперсия:

а) для несгруппированных данных

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n} \quad (26)$$

или

$$\sigma^2 = \bar{x}^2 - (\bar{x})^2 \quad (27)$$

б) для сгруппированных данных

$$\sigma^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f} \quad (28)$$

или

$$\sigma^2 = \bar{x}^2 - (\bar{x})^2 \quad (29)$$

в) для альтернативного признака

$$\sigma_p^2 = g \cdot p \quad (30)$$

4. Среднее квадратическое отклонение:

а) для несгруппированных данных

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \quad (31)$$

б) для сгруппированных данных

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}} \quad (32)$$

в) для альтернативного признака

$$\sigma_p = \sqrt{gp} = \sqrt{p(1-p)} \quad (33)$$

5. Коэффициент вариации:

$$V = \frac{\sigma}{\bar{x}} 100\% \quad (34)$$

6. Внутригрупповая (частная) дисперсия:

$$\text{a)} \quad \sigma_i^2 = \frac{\sum (x - \bar{x}_i)^2}{n} \quad (35)$$

$$\text{б)} \quad \sigma_i^2 = \frac{\sum (x - \bar{x}_i)^2 f}{\sum f}, \quad (36)$$

где X – отдельное значение признака внутри группы;
 $\bar{X}_i$ – средняя арифметическая группы (групповая средняя)

7. Межгрупповая дисперсия:

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 f}{\sum f} \quad (37)$$

Где $\bar{X}_i$ – групповые средние,
 $\bar{X}$ – общая средняя,
 f – численность единиц в группе.

8. Общая средняя внутригрупповых дисперсий:

$$\bar{\sigma}_i^2 = \frac{\sum \sigma_i^2 f}{\sum f} \quad (38)$$

9. Общая дисперсия (согласно правилу сложения дисперсий):

$$\sigma^2 = \bar{\sigma}_i^2 + \delta^2 \quad (39)$$

10. Эмпирический коэффициент детерминации:

$$\eta^2 = \frac{\delta^2}{\sigma^2} \quad (40)$$

при отсутствии связи: $\eta^2 = 0$

при функциональной связи: $\eta^2 = 1$

11. Эмпирическое корреляционное отношение:

$$\eta_{\varepsilon} = \sqrt{\frac{\delta^2}{\sigma^2}} \quad (41)$$

Задача 1

Случайным образом были опрошены 15 сотрудников. При ответе на вопрос о количестве детей в семье были получены следующие данные:

Число детей в семье, чел.	Число ответов
1	8
2	4
3	2
4	1

Определить:

- а) среднее число детей в семье,
- б) моду,
- в) медиану,
- г) дисперсию.

Решение

Расчеты представим в таблице:

X	f	Xf	S	X^2	$X^2 \cdot f$
1	8	8	8	1	8
2	4	8	12	4	16
3	2	6	14	9	18
4	1	4	15	16	16
Итого	15	26	-	30	58

а) $\bar{X} = \frac{\sum Xf}{\sum f} = \frac{26}{15} = 1,7 \text{ детей}$

б) $M_0 = 1$ ребенок

в) $N_{Me} = \frac{\sum f + 1}{2} = \frac{15 + 1}{2} = 8\text{-ой ответ},$
 $Me = 1$ ребенок

г) $\sigma^2 = \bar{x}^2 - (\bar{x})^2 = \frac{\sum x^2 \cdot f}{\sum f} - \left(\frac{\sum xf}{\sum f} \right)^2 = \frac{58}{15} - \left(\frac{26}{15} \right)^2 = 3,87 - 2,89 = 0,98$

Задача 2

На основе данных* о распределении пенсионеров РФ по уровню получаемой пенсии в 1996г.

Пенсия, денном. руб.	Общая численность пенсионеров
	1996г., %
До 150	3,9
150-200	7,9
200-250	17,2
250-300	29,0
300-350	15,4
350-400	12,2
400-500	7,9
Свыше 500	6,5

Определим:

- 1) среднюю пенсию;
- 2) моду;
- 3) медиану, квартили;
- 4) дисперсию;
- 5) среднее квадратическое отклонение;
- 6) коэффициент вариации.

Решение

Расчет представим в следующей таблице.

Пенсия, деном. Руб., X	Численность, %, d	X _{ср}	$\frac{x_{ср}d}{100}$	S	$x_{ср} - \bar{x}$	$(x_{ср} - \bar{x})^2$	$(x_{ср} - \bar{x})^2 \frac{d}{100}$
До 150	3,9	125	4,9	3,9	-180	32400	1263,6
150-200	7,9	175	13,8	11,8	-130	16900	1335,1
200-250	17,2	225	38,7	29,0*	-80	6400	1100,8
250-300	29,0	275	79,8	58,0	-30	900	261
300-350	15,4	325	50,1	73,4	20	400	61,6
350-400	12,2	375	45,8	85,6	70	4900	597,8
400-500	7,9	450	35,6	93,5	145	21025	1661,0
Свыше 500	6,5	550	35,8	100	245	60025	3901,6
Итого	100,0	-	304,5	-	-	-	10182,5

$$1. \bar{x} = \frac{\sum x_{ср} \cdot d}{100} = 304,5 \approx 305 \text{ руб.}$$

2. Модальный интервал: 250-300

$$x_n = 250$$

$$d_2 = 29$$

$$h = 50$$

$$d_3 = 15,4$$

$$d_1 = 17,2$$

$$M_0 = 250 + 50 \frac{29 - 17,2}{(29 - 17,2) + (29 - 15,4)} \approx 273 \text{ руб.}$$

3. Медианный интервал: 250-300

$$x_n = 250$$

$$S_{Me-1}^d = 29^*$$

$$h = 50$$

$$d_{Me} = 29$$

$$N_{Me} = 50$$

$$Me = 250 + 50 \frac{50 - 29^*}{29} \approx 286 \text{ руб.}$$

$$305 > 286 > 273$$

$(\bar{x} > Me > M_o)$, т.е. правосторонняя асимметрия.

$$Q_1 = 200 + 50 \frac{\frac{100}{4} - 11,8}{17,2} \approx 238 \text{ руб.}$$

$$Q_2 = Me$$

$$Q_3 = 350 + 50 \frac{\frac{3}{4} 100 - 73,4}{12,2} \approx 357 \text{ руб.}$$

$$4. \sigma^2 = 10182,5$$

$$5. \sigma = \sqrt{\sigma^2} = \sqrt{10182,5} = 100,9 \text{ руб.}$$

$$6. V = \frac{100,9}{305} \cdot 100\% = 33,08\%$$

Средний размер пенсии в 1996г. составил 305 рублей.
Типичный размер пенсии - 273 рубля.

Половина пенсионеров получали пенсию не более 286 рублей. Другая половина – более этой суммы.

Четверть пенсионеров имеют пенсию не более 238 рублей.
Четверть пенсионеров – более 357 рублей.

Пенсионеры по получаемой пенсии - качественно однородны.

Задача 3

Имеются следующие выборочные данные о вкладах населения района:

Группы населения	Число вкладов, тыс. ед. f	Средний размер вклада, тыс. руб., $\bar{x}_i$	Коэффициент вариаций вклада, %, V
Городское	7	4	20
Сельское	3	6	30
Итого	10	4,6	-

Определим тесноту связи между средним размером вклада и группой населения, исчислив эмпирическое корреляционное отношение.

Решение:

$$\eta_s = \sqrt{\frac{\delta^2}{\sigma^2}}$$

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 f}{\sum f},$$

$$\bar{x} = \frac{\sum x \cdot f}{\sum f} = \frac{4 \cdot 7 + 6 \cdot 3}{7 + 3} = 4,6 \text{ тыс. руб.}$$

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 f}{\sum f} = \frac{(4 - 4,6)^2 \cdot 7 + (6 - 4,6)^2 \cdot 3}{10} = 0,84$$

$$\sigma^2 = \bar{\sigma}_i^2 + \delta^2$$

$$\bar{\sigma}_i^2 = \frac{\sum \sigma_i^2 \cdot f}{\sum f},$$

$$V = \frac{\sigma}{\bar{x}} \cdot 100; \quad \sigma_i = \frac{v_i \cdot \bar{x}}{100}$$

$$\sigma_1 = \frac{20 \cdot 4}{100} = 0,8; \quad \sigma_2 = \frac{30 \cdot 6}{100} = 1,8$$

$$\bar{\sigma}_i^2 = \frac{\sum \sigma_i^2 \cdot f}{\sum f} = \frac{0,8^2 \cdot 7 + 1,8^2 \cdot 3}{7 + 3} = 1,42$$

$$\sigma^2 = 1,42 + 0,84 = 2,26$$

$$\eta^2 = \frac{0,84}{2,26} = 0,3717 \cdot 100 = 37,17\%$$

Т.о. средний размер вкладов только на 37,17% зависит от места проживания и на 62,83% - от внутригрупповых причин.

$$\eta_9 = \sqrt{0,3717} = 0,6097$$

Связь среднего размера вклада и места жительства населения заметная.

Шамшурина Людмила Витальевна

Статистика
(Ряды распределения. Средние величины.)

Учебное издание