Лекция № 7 (продолжение)

Структурные характеристики

Для того чтобы исследовать структуру ряда более подробно, используют структурные показатели, аналогичные медиане. Значения признака, которые делят совокупность на несколько равных по численности групп, получили название *квантилей.* Наиболее часто применяемыми показателями этого типа являются квартили и децили.

Квартили $Q_1,\ Q_2,\ Q_3$ - три значения признака, которые делят совокупность на четыре равных части. Это означает, что ниже первой квартили Q_1 лежат 1/4 часть всех значений x_i и 3/4 значений лежат выше этой величины. Вторая квартиль Q_2 делит распределение пополам и совпадает с медианой. Между медианой и третьей квартилью Q_3 располагается еще 1/4 всей совокупности, и, наконец, 1/4 лежит выше Q_3. Q_1 называется также нижней квартилью, а Q_3 - верхней.

Зная величину квартилей, можно для любого значения ряда определить его положение в ряде. Например, если речь идет о распределении работников предприятия по зарплате, то квартиль Q_3 позволяет определить, попадает ли сотрудник в число 25% наиболее оплачиваемых. Если все ученики класса пробежали кросс и учитель зафиксировал показанное время, то с помощью квартиля Q_1 он сможет определить, входит ли результат данного ученика в число 25% лучших.

Схему нахождения квартилей рассмотрим на конкретном примере.

Пример. Брокер в течение 10 дней фиксировал получаемое им количество заявок и получил следующий результат: 11; 14; 18; 5; 16; 8; 19; 10; 17; 20. Найти значения квартилей распределения.

Решение. Упорядочим ряд по возрастанию: 5; 8; 10; 11; 14; 16; 17; 18; 19; 20. Рассчитаем далее номера квартальных элементов ряда i_1, i_2, i_3:

$$i_1 = N\times\frac{1}{4} = 10\times\frac{1}{4} = 2.5\,; \quad i_2 = N\times\frac{2}{4} = 10\times\frac{2}{4} = 5\,; \quad i_3 = N\times\frac{3}{4} = 10\times\frac{3}{4} = 7.5\,.$$

Если значение индекса дробное, то округлим его до целого и возьмем в качестве квартального значение с полученным округленным индексом; если значение целое, то квартиль будет равен полусумме значений с данным индексом и следующим за ним:

$$i_1 = 2.5 \to 3, \quad Q_1 = x_3 = 10\,;$$

$$i_2 = 5 \to 5, \quad Q_2 = (x_5 + x_6)/2 = (14+16)/2 = 15\,;$$

$$i_3 = 7.5 \to 8, \quad Q_3 = x_8 = 18$$

Для представления квартильной структуры распределения применяют специальный тип графиков, имеющий запоминающееся название "ящик с усами". На рис. 1.10 представлена структура распределения для примера о количестве заявок. Закрашенная область слева и справа ограничена квартилями Q_1, Q_3 линия в центре обозначает медиану, а горизонтальные "усы" показывают интервалы, в которых расположены 25% наименьших и наибольших значений.

Наряду с квартилями используют ***децили*** $(D_1; D_2; \ldots; D_9)$ - значения признака, которые разделяют весь ряд на 10-процентные доли. Децили широко применяются для изучения доходов населения. В этом случае величина D_1 показывает, какова верхняя граница доходов 10% самых бедных людей. Величина же D_9 означает нижнюю границу доходов 10% самых богатых людей. Отношение $\frac{D_9}{D_1}$ представляет собой коэффициент дифференциации - степень расслоения населения по доходам.

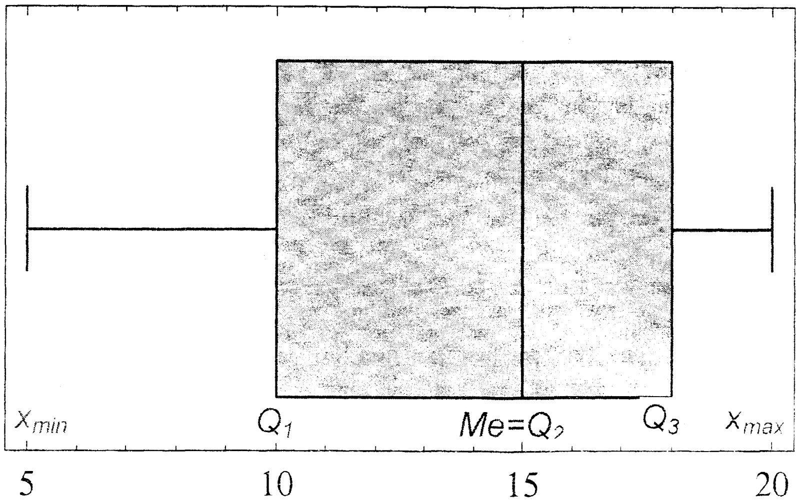

***Рис. 1.10**. Представление структуры распределения на графике "ящик с усами"*

Виды дисперсий. Правило сложения дисперсий

Вариация признака происходит в результате влияния на него различных факторов. Все факторы делятся на основные и второстепенные, поэтому исследователи ставят перед собой задачу расчленения общей вариации результативного признака на вариации под влиянием основного фактора и второстепенных. Эта задача решается с помощью дисперсий.

Отклонение индивидуальных значений результативного признака (y_{ij}) от среднего значения результативного признака для всей совокупности ($\bar{y}$) можно представить как сумму отклонений:

$$y_{ij} - \bar{y} = \left(y_{ij} - \bar{y}_j\right) + \left(\bar{y}_j - \bar{y}\right)$$

где i — текущий номер признака в общей совокупности (от 1 до n);

j — текущий номер группы в интервальном ряду распределения (от 1 до k);

$\bar{y}_j$ — среднее значение результативного признака в j-й группе.

На основании этих отклонений различают следующие виды дисперсии.

Общая дисперсия ($D = \sigma^2$) характеризует вариацию результативного признака под влиянием всех факторов (основного и второстепенных).

$$\sigma^2 = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n}\left(y_{ij} - \overline{y}\right)^2}{n}$$

Межгрупповая дисперсия (δ^2) характеризует вариацию результативного признака под влиянием признака-фактора, положенного в основу группировки.

$$\delta^2 = \frac{\sum_{j=1}^{k}\left(\overline{y}_j - \overline{y}\right)^2 f_j}{\sum_{j=1}^{k} f_j}$$

Внутригрупповые дисперсии (σ_j^2) характеризуют вариацию результативного признака под влиянием неучтенных факторов; рассчитываются для отдельных групп.

$$\sigma_j^2 = \frac{\sum_{i=1}^{f_j}\left(y_{ij} - \overline{y}_j\right)^2}{f_j}$$

Средняя из внутригруппоеых дисперсий ($\overline{\sigma_j^2}$):

$$\overline{\sigma_j^2} = \frac{\sum_{j=1}^{k}\sigma_j^2 f_j}{\sum_{j=1}^{k} f_j}$$

Правило сложения дисперсий: общая дисперсия признака (σ^2) равна сумме межгрупповой дисперсии (δ^2) и средней из внутри-групповых дисперсий ($\overline{\sigma_j^2}$)

$$\sigma^2 = \delta^2 + \overline{\sigma_j^2}$$

Правило сложения дисперсий может быть использовано при определении одной из дисперсий, если известны две другие. Например, зная общую и межгрупповую (факторную) дисперсии, можно определить среднюю дисперсию из внутригрупповых, характеризующую влияние неучтенных факторов.

Использование показателей вариации в анализе взаимосвязей социально-экономических явлений

Для оценки тесноты корреляционной связи между факторным и результативным признаками на базе эмпирического материала строят следующие показатели.

Эмпирический коэффициент детерминации **(**η^2**)** определяется как доля межгрупповой дисперсии в общей дисперсии, характеризует силу влияния факторного (группировочного) признака на результативный.

$$\eta^2 = \frac{\delta^2}{\sigma^2}$$

Напомним, что на величину межгрупповой дисперсии оказывает влияние только факторный признак, а на величину общей дисперсии, помимо факторного признака, и все остальные признаки, поэтому частное от деления межгрупповой дисперсии на общую дисперсию покажет силу влияния факторного признака на результативный.

Эмпирическое корреляционное отношение характеризует тесноту связи; рассчитывается как корень квадратный из эмпирического коэффициента детерминации.

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

Оба показателя находятся в пределах от 0 до 1. При этом чем ближе показатели к единице, тем связь между изучаемыми признаками теснее.

0 — связь отсутствует

1 — связь функциональная.

Для оценки тесноты связи с помощью корреляционного отношения η можно воспользоваться шкалой американского ученого Чеддока:

0,1-0,3 —связь слабая;

0,3-0,5 — связь умеренная;

0,5-0,7 — связь заметная;

0,7-0,9 — связь тесная;

0,9-0,99 — связь весьма тесная.

Дисперсия признака, обладающего альтернативной изменчивостью

При наличии двух взаимоисключающих вариантов значений признака говорят о наличии альтернативной изменчивости признака.

Например, альтернативные значения признака могут быть такими:

— да/нет;

— мужчина/женщина;

— акционер компации/не акционер компании и др.

В этом случае в качестве эквивалента признака можно взять переменную, принимающую только два значения:

1 — когда обследуемая единица совокупности обладает исследуемым признаком;

0 — когда она им не обладает.

Если число единиц, обладающих исследуемым признаком, обозначить f, то число единиц, не обладающих данным признаком, будет равно $n-f$, где n — объем статистической совокупности. Используя данную символику, построим ряд распределения признака, обладающего альтернативной изменчивостью.

Ряд распределения признака, обладающего альтернативной изменчивостью

Варианты признака, x_j	Частота, f_i	Частость, w_j
1 0	f $n-f$	w $1-w$
Итого	n	1

Средняя арифметическая полученного ряда распределения определяется по формуле:

$$\overline{x}_{ар.взв} = \frac{1 \cdot f + 0 \cdot (n - f)}{n} = w$$

Таким образом, среднее значение исследуемого признака, обладающего альтернативной изменчивостью, равно доле исследуемого признака.

Дисперсия исследуемого признака, обладающего альтернативной изменчивостью, определяться по формуле:

$$\sigma^2 = \frac{\sum_{j=1}^{2} \left(x_j - \overline{x}\right)^2 f_j}{\sum_{j=1}^{2} f_j} = \frac{\sum_{j=1}^{2} \left(x_j - \overline{x}\right)^2 w_j}{\sum_{j=1}^{2} w_j} = \frac{(1-w)^2 \cdot w + (0-w)^2 (1-w)}{w + (1-w)} =$$

$$= \left(1 - 2w + w^2\right) w + w^2 (1-w) = w - 2w^2 + w^3 + w^2 - w^3 = w - w^2 = w(1-w)$$

$$\sigma^2 = w(1-w)$$

Таким образом, дисперсия исследуемого признака, обладающего альтернативной изменчивостью, равна произведению доли исследуемого признака на долю альтернативного признака