

Введение. Матрицы как способ представления данных в управлении экономическими системами.

Понятие матрицы и основанный на нем раздел математики – матричная алгебра имеют чрезвычайно важное значение для экономистов. Объясняется это тем, что значительная часть математических моделей экономических объектов и процессов записывается в достаточно простой, а главное – компактной матричной форме. Матрицы позволяют с минимальными затратами труда и времени обрабатывать огромный и весьма разнообразный статистический материал, различные исходные данные, характеризующие уровень, структуру, особенности социально-экономического комплекса.

На данный момент особенно актуально использование матриц для создания баз данных, ведь вся информация обрабатывается и хранится в матричной форме.

Матрица – это прямоугольная таблица, представляющая собой совокупность строк и столбцов. Размерностью матрицы называется величина $m \times n$, где m – число строк, n – число столбцов.

Матрицы обозначают заглавными латинскими буквами A, B, \dots , а их элементы – соответствующими маленькими буквами с двойными индексами, где первый индекс – номер строки, второй – столбца. При записи матрицы таблицу заключают в круглые или квадратные скобки.

Матричный метод позволяет в достаточно простой и понятной форме записывать различные экономические процессы и объекты. Одним из примеров может послужить таблица распределения ресурсов по различным отраслям (табл. 1).

Таблица 1. Распределение ресурсов (в условных единицах)

Ресурсы	Отрасли экономики		
	Промышленность	Сельское хозяйство	Торговля
Трудовые ресурсы	4.8	6.7	7.1
Водные ресурсы	3.1	2.5	5.8
Электрoэнергия	5.6	4.3	3.4

Данная таблица может быть записана в виде матрицы: $A = \begin{pmatrix} 4.8 & 6.7 & 7.1 \\ 3.1 & 2.5 & 5.8 \\ 5.6 & 4.3 & 3.4 \end{pmatrix}$.

Так, например, элемент матрицы $a_{22} = 2.5$ показывает, сколько водных ресурсов потребляет сельское хозяйство, а элемент матрицы $a_{13} = 7.1$ показывает, сколько трудовых ресурсов потребляет торговля.

Диагональными называются элементы матрицы, у которых номер строки совпадает с номером столбца. Эти элементы образуют главную диагональ матрицы, идущую из левого верхнего угла.

Матрица может состоять из одной строки, например, $A = (2 \ 3 \ 4)$ или из одного столбца, например, $B = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$, а также может содержать всего один элемент, например, $C = (3)$.

Матрицы A и B называются равными $A = B$, если у них одинаковые размеры и соответствующие элементы их равны.

Нулевой называют матрицу, все элементы которой равны нулю.

Транспонированной к матрице A называют матрицу A^T , полученную из A заменой каждой строки столбцом с тем же номером, например,

$$A = \begin{pmatrix} 1 & 7 \\ 5 & 0 \end{pmatrix}; \quad A^T = \begin{pmatrix} 1 & 5 \\ 7 & 0 \end{pmatrix}.$$

Квадратной матрицей порядка n называют матрицу, в которой n строк и n

столбцов, например, $B = \begin{pmatrix} 2 & 3 \\ 1 & 5 \end{pmatrix}$ – квадратная матрица второго порядка.

Следом квадратной матрицы A , который обозначают trA , называют сумму элемен-

тов на главной диагонали, например, $A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$, $trA = 1 + 4 = 5$.

Квадратную матрицу называют **диагональной**, если все её элементы,

лежащие вне главной диагонали, равны нулю, например, $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 8 \end{pmatrix}$.

Единичной называют диагональную матрицу, у которой все диагональные элементы равны единице, она обозначается E .

Квадратную матрицу называют **треугольной**, если все её элементы, лежащие по одну сторону главной диагонали, равны нулю. На рисунке 1, где приведены 3 треугольных матрицы, эти элементы окрашены.

$$\begin{pmatrix} 3 & 2 & 5 \\ 0 & 1 & 3 \\ 0 & 0 & 4 \end{pmatrix} \text{ или } \begin{pmatrix} 5 & 0 & 0 \\ 2 & 6 & 0 \\ 3 & 4 & 1 \end{pmatrix} \text{ или } \begin{pmatrix} 1 & 2 & 1 & 3 \\ 0 & 0 & 8 & 4 \\ 0 & 0 & 5 & 7 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Рис.1

Действия над матрицами.

1. Сложение и вычитание существует только для матриц одинакового размера, складывают или вычитают соответствующие элементы матриц. Например, пусть

$$A = \begin{pmatrix} 3 & 12 \\ 1 & 4 \\ 6 & 5 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 8 \\ 0 & 3 \\ 1 & 7 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 2 \\ 3 & 0 \end{pmatrix}.$$

Тогда:

$$A - B = \begin{pmatrix} 3 & 12 \\ 1 & 4 \\ 6 & 5 \end{pmatrix} - \begin{pmatrix} 2 & 8 \\ 0 & 3 \\ 1 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 1 & 1 \\ 5 & -2 \end{pmatrix}, \quad A + B = \begin{pmatrix} 3 & 12 \\ 1 & 4 \\ 6 & 5 \end{pmatrix} + \begin{pmatrix} 2 & 8 \\ 0 & 3 \\ 1 & 7 \end{pmatrix} = \begin{pmatrix} 5 & 20 \\ 1 & 7 \\ 7 & 12 \end{pmatrix}.$$

Матрицы $A \pm C$ и $B \pm C$ не существуют, т.к. A и C , а также B и C – матрицы разного размера.

При умножении или делении матрицы на число все её элементы умножают или делят на это число, например:

$$5 \cdot \begin{pmatrix} 2 & 1 & 3 \\ 4 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 5 & 15 \\ 20 & 0 & 5 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1 & 3 \\ 4 & 0 & 1 \end{pmatrix} : 3 = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 1 \\ \frac{4}{3} & 0 & \frac{1}{3} \end{pmatrix}.$$

Следствие: общий множитель всех элементов матрицы можно выносить за

знак матрицы, например, $\begin{pmatrix} 2 & 8 \\ 6 & 14 \end{pmatrix} = 2 \cdot \begin{pmatrix} 1 & 4 \\ 3 & 7 \end{pmatrix}$.

2. Умножение матриц.

Матрицы можно перемножать только в том случае, когда **число столбцов первой матрицы равно числу строк второй**.

Пусть хотим умножить матрицу A (размера $m \times k$) на матрицу B (размера $k \times n$), т.е. вычислить $C = A \cdot B$. Матрица C имеет размер $m \times n$. Таким образом, $A_{m \times k} \cdot B_{k \times n} = C_{m \times n}$.

Для нахождения элемента c_{ik} матрицы C нужно мысленно поставить вертикально строку № i матрицы A и приложить её к столбцу № k матрицы B , перемножить соответствующие элементы и сложить полученные произведения – см. схему на рисунке 2:

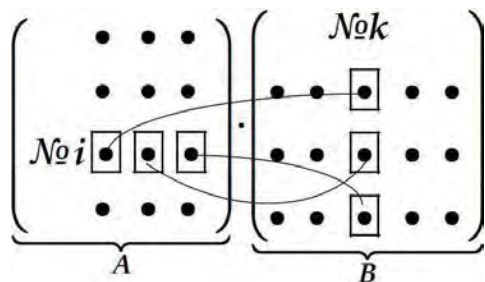


Рис. 2

Пример. Хотим перемножить две матрицы: $B = \begin{pmatrix} 0 & 1 \\ 5 & 1 \end{pmatrix}$ и $A = \begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & 5 \end{pmatrix}$.

Произведение $A \cdot B$ не существует, т.к. у первой матрицы A три столбца, а у второй матрицы B – две строки и $3 \neq 2$. Существует произведение $B \cdot A = C$, т.к. в этом случае у первой матрицы B два столбца, а у второй матрицы A – две строки.

Для вычисления элемента c_{11} разворачиваем вертикально и мысленно прикла-

дываем первую строку матрицы B : $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ к первому столбцу матрицы A : $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$,

т.е. получим $\begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, перемножаем оказавшиеся рядом элементы и результаты

складываем, $c_{11} = 0 \cdot 1 + 1 \cdot 3 = 3$.

Поступая аналогично, находим все остальные элементы матрицы C :

$$\begin{aligned} C &= \begin{pmatrix} 0 & 1 \\ 5 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & 5 \end{pmatrix} = \begin{pmatrix} 0 \cdot 1 + 1 \cdot 3 & 0 \cdot 2 + 1 \cdot 4 & 0 \cdot 0 + 1 \cdot 5 \\ 5 \cdot 1 + 1 \cdot 3 & 5 \cdot 2 + 1 \cdot 4 & 5 \cdot 0 + 1 \cdot 5 \end{pmatrix} = \\ &= \begin{pmatrix} 3 & 4 & 5 \\ 8 & 14 & 5 \end{pmatrix}. \end{aligned}$$

Замечание 1. Если A – квадратная матрица и E – единичная матрица того же размера, то $A \cdot E = E \cdot A$.

Замечание 2. Если существуют оба произведения $A \cdot B$ и $B \cdot A$, то они могут быть матрицами разных размеров.

Например, если матрица A размера 3×4 и матрица B размера 4×3 , то $A \cdot B$ имеет размер 3×3 , а $B \cdot A$ имеет размер 4×4 .

Свойства операций сложения, умножения матриц и умножения их на число α или β :

1. Коммутативность: $A + B = B + A$.

2. Ассоциативность:

$$(A + B) + C = A + (B + C) = A + B + C,$$

$$(A \cdot B) \cdot C = A \cdot (B \cdot C) = A \cdot B \cdot C,$$

$$(\alpha \cdot \beta) \cdot A = \alpha \cdot (\beta \cdot A) = \alpha \cdot \beta \cdot A.$$

3. Дистрибутивность:

$$(A + B) \cdot C = A \cdot C + B \cdot C,$$

$$\alpha \cdot (A + B) = \alpha \cdot A + \alpha \cdot B,$$

$$(\alpha + \beta) \cdot A = \alpha \cdot A + \beta \cdot A.$$

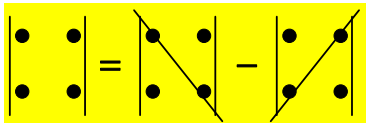
Матрицы широко используются во всех областях науки, в том числе и в экономической науке. Многие обозначения при использовании матриц очень компактны, при этом не теряется ни наглядность, ни содержательность записи. Рассмотрим основные понятия, относящиеся к матрицам.

Определители квадратных матриц

Определителем или детерминантом квадратной матрицы A называется число, обозначаемое $|A|$, либо Δ , либо $\det A$. Его находят по следующим правилам:

1. Определитель матрицы **первого** порядка равен элементу этой матрицы, т.е., если $A = (a_{11})$, то $\Delta = a_{11}$.

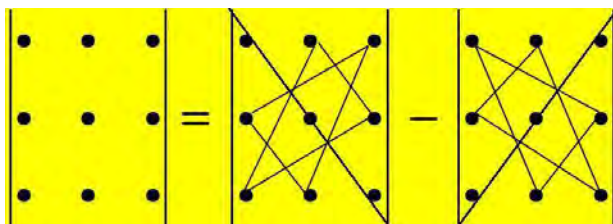
2. Определитель матрицы **второго** порядка $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ вычисляют по

схеме:  , т.е. из произведения элементов, стоящих на

главной диагонали, вычитают произведение элементов, стоящих на по-

бочной диагонали: $\Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} \cdot a_{22} - a_{21} \cdot a_{12}$.

3. Определитель квадратной матрицы **третьего** порядка можно найти по **правилу треугольников**: со знаком «плюс» берут произведения элементов на главной диагонали и произведения элементов, стоящих в вершинах треугольников с основаниями, параллельными ей. Слагаемые со знаком «минус» вычисляют аналогично, но относительно побочной диагонали по схеме:



Пример.
$$\begin{vmatrix} 0 & 2 & 0 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = 0 \cdot 5 \cdot 9 + 2 \cdot 6 \cdot 7 + 4 \cdot 8 \cdot 0 - 7 \cdot 5 \cdot 0 - 4 \cdot 2 \cdot 9 - 8 \cdot 6 \cdot 0 = 12.$$

4. Вычисление определителей высоких порядков.

Рассмотрим квадратную матрицу

$$A_{n \times n} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2k} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ik} & \dots & a_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nk} & \dots & a_{nn} \end{pmatrix}.$$

Минором M_{ik} элемента a_{ik} матрицы A называется определитель матрицы, полученной из A вычёркиванием строки № i и столбца № k .

Алгебраическим дополнением A_{ik} элемента a_{ik} называется его минор, умноженный на $(-1)^{i+k}$.

Т.о. алгебраическое дополнение A_{ik} совпадает с минором M_{ik} , если сумма индексов чётная, и $A_{ik} = -M_{ik}$, если сумма индексов нечётная.

Пример вычисления минора M_{21} и алгебраического дополнения A_{21} элемента a_{21} матрицы A показан на рисунке:

$$A = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}, \quad a_{21} = 2,$$

$$M_{21} = \begin{vmatrix} 4 & 7 \\ 6 & 9 \end{vmatrix} = 4 \cdot 9 - 6 \cdot 7 = -6, \quad A_{21} = -M_{21} = 6.$$

Определитель любого порядка равен сумме произведений элементов любой его строки либо столбца на их алгебраические дополнения.

Это правило называется **разложением определителя** по элементам его строки (столбца). Строки и столбцы называют рядами определителя. Выгодно выбирать ряд, содержащий много нулей.

Пример. Вычислить $\Delta = \begin{vmatrix} 0 & 2 & 0 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix}$.

Разложим определитель по элементам первой строки, т.к. в ней два нуля:

$$\begin{aligned} \Delta &= 0 \cdot A_{11} + 2 \cdot A_{12} + 0 \cdot A_{13} = -2 \cdot M_{12} = \\ &= -2 \cdot \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} = -2 \cdot (4 \cdot 9 - 7 \cdot 6) = -2 \cdot (36 - 42) = 12. \end{aligned}$$

Ранее этот же ответ был получен по правилу треугольников.

Свойства определителей

1. Определитель не меняется при замене строк столбцами с тем же номером, т.е. определители исходной и транспонированной матриц совпадают.
2. Определитель изменяет знак при перестановке двух параллельных рядов.
3. Определитель, имеющий 2 одинаковых параллельных ряда, равен нулю.
4. Определитель равен нулю, если любой его ряд состоит из нулей.
5. Общий множитель элементов любого ряда можно вынести за знак определителя, например, на схеме внизу выделены ряды, из которых вынесен общий множитель:

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 3 \\ 4 & 8 & 6 \\ 7 & 8 & 9 \end{vmatrix} &= 2 \cdot \begin{vmatrix} 1 & 2 & 3 \\ \boxed{2} & \boxed{4} & \boxed{3} \\ 7 & 8 & 9 \end{vmatrix} = 2 \cdot 3 \cdot \begin{vmatrix} 1 & 2 & \boxed{1} \\ 2 & 4 & \boxed{1} \\ 7 & 8 & \boxed{3} \end{vmatrix} = \underbrace{2 \cdot 3 \cdot 2}_{12} \cdot \begin{vmatrix} 1 & \boxed{1} & 1 \\ 2 & \boxed{2} & 1 \\ 7 & \boxed{4} & 3 \end{vmatrix} = \\ &= 12 \cdot (1 \cdot 2 \cdot 3 + 2 \cdot 4 \cdot 1 + 1 \cdot 1 \cdot 7 - 7 \cdot 2 \cdot 1 - 2 \cdot 1 \cdot 3 - 4 \cdot 1 \cdot 1) = -36. \end{aligned}$$

6. Определитель не изменится, если к элементам одного ряда прибавить соответствующие элементы параллельного ряда, умноженные на любое

число например: $\begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} = 4 - 6 = -2$. Прибавим к элементам первой строки

соответствующие элементы второй, умноженные на 2, и найдём новый опре-

делитель: $\begin{vmatrix} 1+2 \cdot 2 & 3+4 \cdot 2 \\ 2 & 4 \end{vmatrix} = \begin{vmatrix} 5 & 11 \\ 2 & 4 \end{vmatrix} = 20 - 22 = -2$. Ответы совпали.

7. Если элементы любого ряда определителя являются суммой двух слагаемых, то определитель равен сумме соответствующих определителей:

$$\begin{vmatrix} a & c+d \\ b & g+h \end{vmatrix} = \begin{vmatrix} a & c \\ b & g \end{vmatrix} + \begin{vmatrix} a & d \\ b & h \end{vmatrix}.$$

8. Если A и B – матрицы одинакового порядка, то **определитель произведения** этих матриц **равен произведению их определителей**, т.е.

$$|A \cdot B| = |B \cdot A| = |A| \cdot |B|,$$

даже если $A \cdot B \neq B \cdot A$.

9. Определитель **треугольной** или **диагональной** матрицы равен произведению её диагональных членов, например,

$$\begin{vmatrix} 1 & 254 & 348 \\ 0 & 2 & 576 \\ 0 & 0 & 3 \end{vmatrix} = 1 \cdot 2 \cdot 3 = 6, \quad \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{vmatrix} = 1 \cdot 2 \cdot 3 \cdot 4 = 24.$$

Обратная матрица

Рассмотрим квадратную матрицу n -ого порядка. Она называется невырожденной, если её определитель $\Delta = |A| \neq 0$. Для невырожденной матрицы A существует обратная матрица, обозначаемая A^{-1} . Матрица A^{-1} называется обратной матрице A , если выполняются условия $A \cdot A^{-1} = A^{-1} \cdot A = E$, где E – единичная матрица того же порядка.

Рассмотрим методы нахождения обратной матрицы.

1. Метод присоединённой матрицы

Составим матрицу \tilde{A} из алгебраических дополнений A_{ik} элементов a_{ik} исходной

матрицы A :
$$\tilde{A} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{pmatrix}.$$

Транспонированная матрица \tilde{A}^T называется присоединённой матрицей. Тогда

$$A^{-1} = \frac{1}{\Delta} \cdot \tilde{A}^T = \frac{1}{\Delta} \cdot \begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{pmatrix}.$$

Пример. $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Найти A^{-1} , если она существует.

Решение. $\Delta = \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 1 \cdot 4 - 2 \cdot 3 = -2 \neq 0$, следовательно, A^{-1} существует. Находим алгебраические дополнения элементов матрицы:

$$\begin{aligned} A_{11} &= M_{11} = 4, & A_{21} &= -M_{21} = -2, \\ A_{12} &= -M_{12} = -3, & A_{22} &= M_{22} = 1. \end{aligned}$$

$$A^{-1} = \frac{1}{\Delta} \cdot \begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix} = \frac{1}{-2} \cdot \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}.$$

2. Метод Гаусса

Составляем расширенную матрицу, приписывая к матрице A единичную матрицу E того же порядка. Приписанную матрицу будем отделять вертикальной чертой от основной матрицы.

Затем, используя элементарные преобразования строк матрицы (перестановка, умножение всех элементов строки на число $\neq 0$, добавление ко всем элементам строки соответствующих элементов другой строки, умноженных на одно и то же число), добиваемся, чтобы на месте матрицы A возникла единичная матрица E .

Тогда на месте E автоматически возникает A^{-1} .

Матрицы A и B называются эквивалентными и обозначаются $A \sim B$, если одна получена из другой элементарными преобразованиями её рядов.

Пример. Рассмотрим снова матрицу $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Вычислим A^{-1} методом Гаусса.

Составляем расширенную матрицу. Справа от матрицы записываем действия, производимые с её строками (строки нумеруем римскими цифрами):

$$\begin{aligned}
& \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array} \right) II - I \cdot 3 \sim \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array} \right) II : (-2) \sim \\
& \sim \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array} \right) I - II \cdot 2 \sim \left(\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array} \right); \quad A^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}, \\
& \quad \quad \quad \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_E \quad \underbrace{\begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}}_{A^{-1}}
\end{aligned}$$

что совпадает с предыдущим результатом.

Пример 1. Решить систему $\begin{cases} 3x + 2y - 6z = -11 \\ 2x - 3y + 5z = 11 \\ x + y + z = 6 \end{cases}$ по формулам Крамера.

Решение. По правилу треугольников находим:

$$\Delta = \begin{vmatrix} 3 & 2 & -6 \\ 2 & -3 & 5 \\ 1 & 1 & 1 \end{vmatrix} = -9 + 10 - 12 - 18 - 15 - 4 = -48,$$

$$\Delta_1 = \begin{vmatrix} -11 & 2 & -6 \\ 11 & -3 & 5 \\ 6 & 1 & 1 \end{vmatrix} = 33 + 60 - 66 - 108 - 22 + 55 = -48,$$

$$\Delta_2 = \begin{vmatrix} 3 & -11 & -6 \\ 2 & 11 & 5 \\ 1 & 6 & 1 \end{vmatrix} = 33 - 55 - 72 + 66 + 22 - 90 = -96,$$

$$\Delta_3 = \begin{vmatrix} 3 & 2 & -11 \\ 2 & -3 & 11 \\ 1 & 1 & 6 \end{vmatrix} = -54 + 22 - 22 - 33 - 24 - 33 = -144,$$

$$x = \frac{\Delta_1}{\Delta} = \frac{-48}{-48} = 1, \quad y = \frac{\Delta_2}{\Delta} = \frac{-96}{-48} = 2, \quad z = \frac{\Delta_3}{\Delta} = \frac{-144}{-48} = 3.$$

Пример 2. Рассмотрим и решим с помощью системы линейных уравнений следующую задачу: из определенного листового материала необходимо выкроить 360 заготовок типа А, 300 заготовок типа Б и 675 заготовок типа В. При этом можно применять три способа раскроя. Количество заготовок, получаемых из каждого листа при каждом способе раскроя, указано в таблице:

Тип заготовки	Способ раскроя		
	1	2	3
А	3	2	1
Б	1	6	2
В	4	1	5

Требуется найти сколько листов материала нужно для каждого способа раскроя.

Решение. Запишем в математической форме условия выполнения задания.

Обозначим через x, y, z количество листов материала, раскраиваемых соответственно первым, вторым и третьим способами. Тогда при первом способе раскроя x листов будет получено $3x$ заготовок типа А, при втором – $2y$, при третьем – z .

Для полного выполнения задания по заготовкам типа А должно выполняться равенство: $3x + 2y + z = 360$.

Аналогично для второго и третьего способов раскроя получаем уравнения:

$x + 6y + 2z = 300$, $4x + y + 5z = 675$. Решаем систему уравнений

$$\begin{cases} 3x + 2y + z = 360, \\ x + 6y + 2z = 300, \\ 4x + y + 5z = 675. \end{cases}$$

Полученная система уравнений выражает в математической форме условие выполнения всего задания по заготовкам А, Б и В. Решаем систему по формулам Крамера:

$$\Delta = \begin{vmatrix} 3 & 2 & 1 \\ 1 & 6 & 2 \\ 4 & 1 & 5 \end{vmatrix} = 67, \quad \Delta_1 = \begin{vmatrix} 360 & 2 & 1 \\ 300 & 6 & 2 \\ 675 & 1 & 5 \end{vmatrix} = 6030, \quad x = \frac{\Delta_1}{\Delta} = \frac{6030}{67} = 90,$$

$$\Delta_2 = \begin{vmatrix} 3 & 360 & 1 \\ 1 & 300 & 2 \\ 4 & 675 & 5 \end{vmatrix} = 1005, \quad y = \frac{\Delta_2}{\Delta} = \frac{1005}{67} = 15,$$

$$\Delta_3 = \begin{vmatrix} 3 & 2 & 360 \\ 1 & 6 & 300 \\ 4 & 1 & 675 \end{vmatrix} = 4020, \quad z = \frac{\Delta_3}{\Delta} = \frac{4020}{67} = 60.$$

Ответ. $x = 90$, $y = 15$, $z = 60$.

Решение систем линейных уравнений методом обратной матрицы

Так можно решать лишь системы, в которых **число уравнений совпадает с числом неизвестных**. Рассмотрим систему (1) и предположим, что её определитель $\Delta \neq 0$, следовательно, обратная матрица A^{-1} существует. Запишем равенство (1) в матричной форме

$$A \cdot X = B \quad (3)$$

и умножим слева обе части на A^{-1} :

$$A^{-1}(AX) = A^{-1}B. \quad (4)$$

Поскольку

$$A^{-1}(AX) = (A^{-1}A)X = EX = X, \quad (5)$$

то из (4) и (5) получим

$$X = A^{-1} \cdot B. \quad (6)$$

Пример. Методом обратной матрицы решить систему
$$\begin{cases} x_1 + 2x_2 = 5, \\ 3x_1 + 4x_2 = 11. \end{cases}$$

Решение. $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 5 \\ 11 \end{pmatrix}$, $\Delta = \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 4 - 6 = -2$.

Найдём сначала обратную матрицу методом присоединённой матрицы.

Вычисляем алгебраические дополнения A_{ik} элементов a_{ik} матрицы A .

$$A_{11} = 4, \quad A_{12} = -3, \quad A_{21} = -2, \quad A_{22} = 1.$$

$$A^{-1} = \frac{1}{\Delta} \cdot \begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix} = \frac{1}{-2} \cdot \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}.$$

По формуле (6) находим:

$$X = A^{-1} \cdot B = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} 5 \\ 11 \end{pmatrix} = \begin{pmatrix} (-2) \cdot 5 + 1 \cdot 11 \\ \frac{3}{2} \cdot 5 - \frac{1}{2} \cdot 11 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Получили: $x_1 = 1$; $x_2 = 2$.

Сначала рассмотрим вспомогательные понятия. Назовём строки и столбцы рядами матрицы. Из матрицы A размера $m \times n$ можно получить вычёркиванием произвольных рядов квадратные матрицы размера $k \times k$, где $k \leq \min(m, n)$.

Определители таких квадратных матриц называют **минорами** порядка k и обозначают M_k . Например из матрицы A размера 3×4 можно получить квадратные матрицы размера 1×1 , 2×2 , 3×3 . Их определители будут соответственно минорами I, II и III порядка.

Пример. Рассмотрим матрицу $A = \begin{pmatrix} 2 & 1 & 4 & 2 \\ 3 & 5 & 7 & 1 \\ 9 & 8 & 0 & 6 \end{pmatrix}$ и найдём её миноры.

- Миноры I порядка – все элементы матрицы: $M_1 = 2$, $M_1 = 3$ и т.д.
- Миноры II порядка получим, вычёркивая из A любую строку и 2 столбца, например, вычеркнули III строку и III и IV столбцы получили минор

$$M_2 = \begin{vmatrix} 2 & 1 \\ 3 & 5 \end{vmatrix} = 7.$$

При вычёркивании II строки и I и III столбцов получим минор

$$M_2 = \begin{vmatrix} 1 & 2 \\ 8 & 6 \end{vmatrix} = -10 \quad \text{и т.д.}$$

- Миноры III порядка получим, вычёркивая из A любой столбец, например,

$$\text{второй: } M_3 = \begin{vmatrix} 2 & 4 & 2 \\ 3 & 7 & 1 \\ 9 & 0 & 6 \end{vmatrix} = 84 + 36 - 126 - 72 = -78.$$

Рангом матрицы A называют наивысший порядок её отличных от нуля миноров.

Ранг обозначают $r(A)$ или $\text{rang}(A)$. Из определения следует:

1. Ранг не превосходит меньшего из размеров матрицы: $r(A) \leq \min(m, n)$;
2. $r(A) = 0$ только, когда все элементы матрицы равны нулю;
3. $r(A) = n$ для квадратной матрицы порядка n , если её определитель $\Delta \neq 0$.

Базисным называют каждый отличный от нуля минор, порядок которого равен рангу матрицы. Таких миноров может быть несколько.

Пример. Найти ранг матрицы: $A = \begin{pmatrix} 2 & 0 & 4 & 0 \\ \boxed{3} & 0 & \boxed{6} & 0 \\ \boxed{1} & 0 & \boxed{-3} & 0 \end{pmatrix}$.

Решение. Поскольку в данной матрице 3 строки и 4 столбца, её ранг не превзойдёт меньшего из этих чисел, т.е. $r(A) \leq 3$.

Все миноры третьего порядка равны нулю, т.к. содержат нулевой столбец. Имеется отличный от нуля минор II порядка, его элементы выделены в матрице.

$M_2 = \begin{vmatrix} 3 & 6 \\ 1 & -3 \end{vmatrix} = -15 \neq 0$. Следовательно ранг матрицы $r(A) = 2$. Этот минор – ба-

зисный. Есть ещё базисный минор $M_2 = \begin{vmatrix} 2 & 4 \\ 1 & -3 \end{vmatrix} = -10 \neq 0$.

Не является базисным минор $M_2 = \begin{vmatrix} 2 & 4 \\ 3 & 6 \end{vmatrix} = 0$.

Находить ранг матрицы перебором всех её миноров долго. Для облегчения задачи используют преобразования матрицы, сохраняющие её ранг:

1. **Отбрасывание нулевого ряда.**
2. **Умножение всех элементов ряда на число $\neq 0$.**
3. **Перестановка двух параллельных рядов.**
4. **Транспонирование матрицы.**
5. **Добавление к каждому элементу ряда соответствующих элементов параллельного ряда, умноженных на любое число.**

Матрицы A и B называют эквивалентными и обозначают $A \sim B$, если одна получается из другой этими преобразованиями.

Ступенчатой называют матрицу, у которой элементы на главной диагонали отличны от нуля, а под диагональю – нули. Нули получают последовательно в I , II , и т.д. столбцах, двигаясь слева направо. Такая матрица схематично

изображена на рисунке:

$$\begin{pmatrix} \begin{matrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot \end{matrix} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

(Note: The diagram shows a \$6 \times 6\$ matrix. The first \$3 \times 3\$ submatrix is highlighted with a purple box and labeled \$k\$ on both the left and bottom. The diagonal elements of this submatrix are 0, 0, 0. The rest of the matrix contains dots representing non-zero elements.)

Способ нахождения базисного минора

Нужно, используя преобразования, сохраняющие ранг матрицы, привести её к ступенчатому виду. Если в матрице останется k строк, то её ранг равен k , а первые k столбцов слева образуют базисный минор (он обведён на рисунке).

Пример. Дана матрица: $A = \begin{pmatrix} 0 & 4 & 10 & 1 \\ 4 & 0 & -2 & 5 \\ 2 & 2 & 4 & 3 \\ 1 & 3 & 7 & 2 \end{pmatrix}$. Найти её ранг и базисный минор.

Решение. Переместим вверх последнюю строку, чтобы получить для удобства единицу в верхнем левом углу матрицы. Будем получать под главной диагональю нули последовательно в первом, втором и т.д. столбцах и воспользуемся преобразованиями, не меняющими ранг матрицы. Справа от матрицы записываем действия, проводимые с соответствующими строками.

$$\begin{aligned} A &\sim \begin{pmatrix} 1 & 3 & 7 & 2 \\ 0 & 4 & 10 & 1 \\ 4 & 0 & -2 & 5 \\ 2 & 2 & 4 & 3 \end{pmatrix} \begin{matrix} \\ III - IV \cdot 2 \\ IV - I \cdot 2 \end{matrix} \sim \begin{pmatrix} 1 & 3 & 7 & 2 \\ 0 & 4 & 10 & 1 \\ 0 & -4 & -10 & -1 \\ 0 & -4 & -10 & -1 \end{pmatrix} \begin{matrix} \\ \\ III + II \\ IV - III \end{matrix} \sim \\ &\sim \begin{pmatrix} 1 & 3 & 7 & 2 \\ 0 & 4 & 10 & 1 \\ \emptyset & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset \end{pmatrix} \sim \left(\begin{array}{cc|cc} 1 & 3 & 7 & 2 \\ 0 & 4 & 10 & 1 \end{array} \right). \end{aligned}$$

Т.к. осталось 2 строки, то $r(A) = 2$ и слева автоматически получился базисный

минор $M_2 = \begin{vmatrix} 1 & 3 \\ 0 & 4 \end{vmatrix} = 4 \neq 0$, который обведён. Имеются ещё базисные миноры

$$M_2 = \begin{vmatrix} 1 & 7 \\ 0 & 10 \end{vmatrix} = 10, \quad M_2 = \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} = 1.$$

Решение системы линейных уравнений методом Гаусса.

Теорема Кронекера-Капелли

Рассмотрим систему m линейных уравнений с n неизвестными:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases} \quad (1)$$

В матричной форме такая система записывается в виде $A \cdot X = B$, где

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}; \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}; \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}. \quad (2)$$

Матрица A называется матрицей системы, X – столбец неизвестных, B – столбец свободных членов.

Запишем расширенную матрицу \bar{A} , приписав к A справа столбец свободных членов B , который отделяем вертикальной чертой от основной матрицы.

$$\bar{A} = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right) \quad (3)$$

Теорема 1 (Кронекера-Капелли). Система (1) совместна, т.е. имеет хотя бы одно решение, если ранг основной матрицы A совпадает с рангом расширенной \bar{A} , т.е. $r(A) = r(\bar{A})$, в противном случае решений нет.

Теорема 2. Система (1) имеет единственное решение, если ранг матрицы A равен числу неизвестных.

Отметим, что для квадратной матрицы A это равносильно тому, что определитель матрицы не должен равняться нулю.

Теорема 3. Система (1) имеет бесконечное множество решений, если ранг матрицы A **меньше** числа неизвестных.

Для решения системы линейных уравнений **методом Гаусса** разбиваем требуемые действия на 4 этапа.

1. Одновременно вычисляем ранги матриц A и \bar{A} . Для этого приводим расширенную матрицу \bar{A} к ступенчатому виду, преобразуя **только её строки**. Если $r(A) < r(\bar{A})$, то система несовместна, решений нет.
2. Если $r(A) = r(\bar{A}) = n$, то система совместна и имеет единственное решение, которое можно найти любым способом. Если $r(A) = r(\bar{A}) < n$ система имеет бесконечное множество решений. Находим базисный минор и оставляем только те уравнения, из которых он составлен, остальные отбрасываем.

Главными или **базисными** называем неизвестные, **коэффициенты при которых входят в базисный минор**, их оставляем слева. Остальные неизвестные называем

свободными и переносим в правые части уравнений. Обозначаем свободные неизвестные греческими буквами α, β, γ и т.д.

3. Получаем общее решение системы, выражая главные неизвестные через свободные.
4. Находим частные решения системы, присваивая свободным неизвестным произвольные значения.

Замечание. Поскольку при приведении матрицы к ступенчатому виду мы последовательно исключаем неизвестные (получая нули в соответствующих местах), метод Гаусса называют также **методом последовательного исключения неизвестных**. Для системы с квадратной матрицей, определитель которой отличен от нуля, матрицу A преобразуют в единичную. При этом получение нулей под главной диагональю называется прямым ходом метода Гаусса, а над ней – обратным ходом.

Пример. Исследовать и решить систему:
$$\begin{cases} 2x_1 + 7x_2 + 3x_3 + x_4 = 6 \\ 3x_1 + 5x_2 + 2x_3 + 2x_4 = 4 \\ 9x_1 + 4x_2 + x_3 + 7x_4 = 2 \end{cases}$$

Решение. Составляем расширенную матрицу и приводим её к ступенчатому виду:

$$\begin{aligned} \bar{A} &= \left(\begin{array}{cccc|c} 2 & 7 & 3 & 1 & 6 \\ 3 & 5 & 2 & 2 & 4 \\ 9 & 4 & 1 & 7 & 2 \end{array} \right) \xrightarrow{II \cdot 2 - I \cdot 3, III - II \cdot 3} \left(\begin{array}{cccc|c} 2 & 7 & 3 & 1 & 6 \\ 0 & -11 & -5 & 1 & -10 \\ 0 & -11 & -5 & 1 & -10 \end{array} \right) \sim \\ &\sim \left(\begin{array}{cccc|c} 2 & 7 & 3 & 1 & 6 \\ 0 & -11 & -5 & 1 & -10 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right) \sim \left(\begin{array}{cccc|c} \boxed{2} & 7 & 3 & \boxed{1} & 6 \\ \boxed{0} & -11 & -5 & \boxed{1} & -10 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right). \end{aligned} \quad (1)$$

Слева автоматически получился общий для A и \bar{A} базисный минор

$$M_2 = \begin{vmatrix} 2 & 7 \\ 0 & -11 \end{vmatrix} = -22, \quad r(A) = r(\bar{A}) = 2. \text{ Для простоты расчётов удобнее выбрать}$$

другой общий базисный минор $M_2 = \begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} = 2$, его элементы обведены в формуле

(1). Берём уравнения, из которых он составлен.
$$\begin{cases} 2x_1 + 7x_2 + 3x_3 + x_4 = 6, \\ 0 \cdot x_1 - 11x_2 - 5x_3 + x_4 = -10. \end{cases}$$

Главные переменные x_1 и x_4 оставляем слева, свободные x_2 и x_3 переносим вправо:

$$\begin{cases} 2x_1 + x_4 = -7x_2 - 3x_3 + 6, \end{cases} \quad (2)$$

$$\begin{cases} 0 \cdot x_1 + x_4 = 11x_2 + 5x_3 - 10. \end{cases} \quad (3)$$

Присвоим x_2 и x_3 произвольные значения: $x_2 = \alpha$, $x_3 = \beta$. Найдём из уравнения (3): $x_4 = 11\alpha + 5\beta - 10$.

Из уравнения (2):

$$x_1 = \frac{-7x_2 - 3x_3 - x_4 + 6}{2} = \frac{-7\alpha - 3\beta - 11\alpha - 5\beta + 10 + 6}{2} = -9\alpha - 4\beta + 8.$$

Ответ: система имеет бесчисленное множество решений.

Общее решение:
$$\begin{cases} x_1 = -9\alpha - 4\beta + 8, \\ x_2 = \alpha, \\ x_3 = \beta, \\ x_4 = 11\alpha + 5\beta - 10. \end{cases} \quad (4)$$

Частное решение: $\alpha = 1, \beta = 1, x_1 = -5, x_2 = 1, x_3 = 1, x_4 = 6$.

Замечание. Необходимо сделать проверку и подставить общее решение **во все уравнения исходной системы**. Привести подобные члены и убедиться, что правые части уравнений совпали.

Проверка. Подставляем формулы (4) в исходную систему уравнений.

Получим:
$$\begin{cases} 2(-9\alpha - 4\beta + 8) + 7\alpha + 3\beta + 11\alpha + 5\beta - 10 = 6, \\ 3(-9\alpha - 4\beta + 8) + 5\alpha + 2\beta + 2(11\alpha + 5\beta - 10) = 4, \\ 9(-9\alpha - 4\beta + 8) + 4\alpha + \beta + 7(11\alpha + 5\beta - 10) = 2. \end{cases}$$

Правые части совпали с исходными, решение верное.

Исследования показывают, что в окружающем нас мире величины тесно связаны друг с другом, например, цена товара и величина спроса на него, инфляция и безработица, объём производства и прибыль и т.д. Поэтому функции широко применяются в экономике, например, функция полезности, производственная функция, функция выпуска, функция издержек, функция спроса.

Рассмотрим основные понятия, связанные с функциями.

Числовые промежутки, отрезки, интервалы

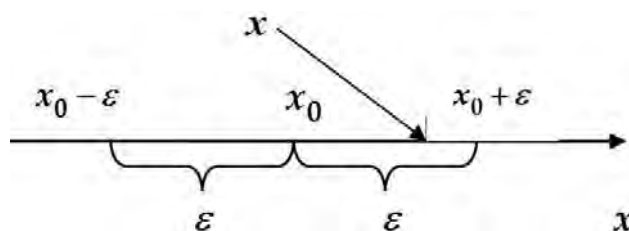
Пусть a и b — действительные числа.

1. Множество чисел x , удовлетворяющих неравенствам $a < x < b$, называется интервалом (a, b) .
2. Множество чисел x , удовлетворяющих неравенствам $a \leq x \leq b$, называется отрезком $[a, b]$.
3. Множество чисел x , удовлетворяющих неравенствам $a \leq x < b$ либо $a < x \leq b$, называется соответственно полуинтервалом $[a, b)$ либо $(a, b]$.
4. Интервалы, отрезки и полуинтервалы называют также промежутками.

Окрестностью точки x_0 называется любой интервал, содержащий эту точку.

Интервал $(x_0 - \varepsilon, x_0 + \varepsilon)$, где $\varepsilon > 0$, называется ε -окрестностью точки x_0 .

Если $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$, то точка x попадает в ε -окрестность точки x_0 .



Тогда $x_0 - \varepsilon < x < x_0 + \varepsilon$, $-\varepsilon < x - x_0 < \varepsilon$ или $|x - x_0| < \varepsilon$. Выполнение неравенства $|x - x_0| < \varepsilon$ означает попадание точки x в ε -окрестность точки x_0 .

Функция

Функцией $y = f(x)$ называется правило, по которому каждому элементу x

множества X ставится в соответствие единственный элемент y множества Y .

Говорят, что функция $y = f(x)$ задана на множестве X , “ x ” называется аргументом или независимой переменной, “ y ” называется зависимой переменной или функцией, буква “ f ” обозначает закон соответствия.

Если элементы множеств X и Y – действительные числа, то функцию называют числовой. Множество X называется **областью определения** функции, а множество Y – **областью значений** функции. Обычно под областью определения функции подразумевается область допустимых значений x , т.е. таких x , при которых функция $y = f(x)$ имеет смысл. Например, область определения функции $y = \sqrt{1-x^2}$ – отрезок $[-1; 1]$.

Для задания функции нужно **указать правило**, позволяющее находить y , зная x .

Пример. Дана функция $y(x) = \frac{1+x}{1-x}$. Найти $y\left(\frac{4-x}{2+x}\right)$.

Решение. Нужно в формуле (1) заменить x выражением $\frac{4-x}{2+x}$. Получим:

$$y\left(\frac{4-x}{2+x}\right) = \frac{1 + \frac{4-x}{2+x}}{1 - \frac{4-x}{2+x}} = \frac{2+x+4-x}{2+x-4+x} = \frac{6}{2x-2} = \frac{3}{x-1}.$$

Абсолютной величиной или модулем числа x называется само число x , если

$$x \geq 0 \text{ и число “} -x \text{”, если } x < 0: |x| = \begin{cases} x, & \text{если } x \geq 0; \\ -x, & \text{если } x < 0. \end{cases}$$

Свойства абсолютной величины:

1. $|x| \geq 0$.
2. $|x| = |-x|$.
3. $-|x| \leq x \leq |x|$.
4. $|x \cdot y| = |x| \cdot |y|$.
5. $\left|\frac{x}{y}\right| = \frac{|x|}{|y|}$, если $y \neq 0$.
6. Если $\varepsilon > 0$, то неравенства $|x| \leq \varepsilon$ и $-\varepsilon \leq x \leq \varepsilon$ равносильны.
7. Неравенство $|x| > \varepsilon$ и пара неравенств $\begin{cases} x > \varepsilon \\ x < -\varepsilon \end{cases}$ равносильны.

Пример. Решить неравенство: $|x+3| < 5$.

Решение. Из свойства (6) находим: $-5 < x + 3 < 5$, $-8 < x < 2$.

Способы задания функции

1. Аналитический, при котором функция задаётся формулой вида $y = f(x)$, например, $y = x^2 + \sqrt{10 - x}$.
2. Табличный, при котором функция задаётся таблицей, содержащей значения аргумента x и соответствующие значения $f(x)$, например, таблица синусов.
3. Графический, при котором функция задаётся графиком, т.е. множеством точек (x, y) плоскости, где x – значения аргумента, y – соответствующие значения функции $y = f(x)$.

Основные свойства функций

1. Функция $y = f(x)$, область определения которой симметрична относительно нуля, называется **чётной**, если для любых значений x из области её определения $f(-x) = f(x)$ и называется **нечётной**, если $f(-x) = -f(x)$. График чётной функции симметричен относительно оси Oy т.к., по определению, вместе с любой точкой (x, y) он содержит точку $(-x, y)$ – рис.1.

График нечётной функции симметричен относительно начала координат, т.к., по определению, вместе с любой точкой (x, y) он содержит точку $(-x, -y)$ – рис. 2.

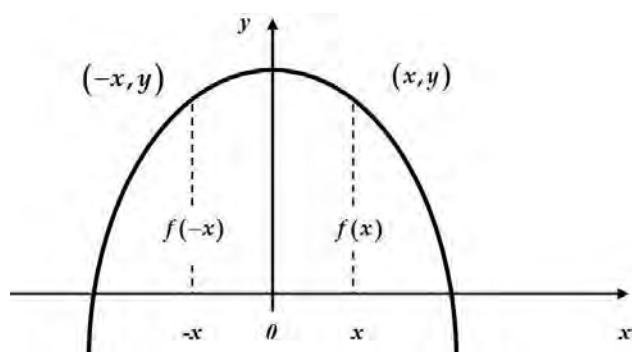


Рис. 1

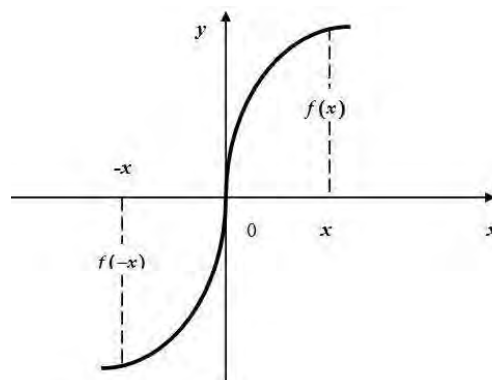


Рис. 2

Функция, не являющаяся ни чётной, ни нечётной, называется функцией **общего вида**.

Примеры. Определить тип функции.

$$1. f(x) = \frac{x^3}{x^2 + 1}, f(-x) = \frac{(-x)^3}{(-x)^2 + 1} = -\frac{x^3}{x^2 + 1} = -f(x) - \text{нечётная.}$$

$$2. f(x) = x^4 - 5|x|, f(-x) = (-x)^4 - 5|-x| = x^4 - 5|x| = f(x) - \text{чётная.}$$

$$3. f(x) = e^x - 2e^{-x}, f(-x) = e^{-x} - 2e^x \neq f(x) \neq -f(x) - \text{общего вида.}$$

Функция $y = f(x)$ называется **возрастающей**, если для любых двух значений аргумента $x_2 > x_1$ из области её определения выполняется неравенство $f(x_2) > f(x_1)$.

Функция $y = f(x)$ называется **убывающей**, если для любых двух значений аргумента $x_2 > x_1$ из области её определения выполняется неравенство $f(x_2) < f(x_1)$.

Возрастающие и убывающие функции называются **строго монотонными**. Если на отдельных участках области определения функция может оставаться постоянной, т.е. для $x_2 > x_1$ выполняются нестрогие неравенства $f(x_2) \geq f(x_1)$ или $f(x_2) \leq f(x_1)$, то функция называется **монотонной**, в I случае **неубывающей**, во II - **невозрастающей**.

Интервалы, на которых функция монотонна, называются интервалами монотонности.

Например, функция $y = x^2$ при $x \in (-\infty, 0]$ убывает, при $x \in [0, \infty)$ – возрастает.

- Функция $y = f(x)$ называется **ограниченной** на промежутке X , если существует такое положительное число $M > 0$, что $|f(x)| \leq M \quad \forall x \in X$. В противном случае функция называется неограниченной. Например, функция $y = \sin x$ ограничена на всей числовой оси, т.к. $|\sin x| \leq 1$ для любого действительного x .

- Функция $y = f(x)$ называется **периодической**, если $\forall x \in X$:

$$f(x + T) = f(x), \quad (1)$$

где T – постоянное число. Наименьшее положительное число T , удовлетворяющее этому условию, называется **периодом** функции. Например, функция $y = \sin x$ имеет период $T = 2\pi$.

Если T – период, то периодами будут и числа $T \cdot n$, где $n = \pm 1, \pm 2, \dots$ Поэтому под словом «период» понимают наименьшее положительное число, удовлетворяющее равенству (1).

Пример. Найти период функции $f(x) = \sin 4x$, если он существует.

Решение. По формуле (1) находим:

$$\sin(4(x+T)) = \sin 4x, \quad \sin(4x+4T) - \sin 4x = 0. \text{ Учитывая, что}$$

$$\sin \alpha - \sin \beta = 2 \sin \frac{\alpha - \beta}{2} \cdot \cos \frac{\alpha + \beta}{2}, \text{ получим:}$$

$$\sin(4x+4T) - \sin 4x = 2 \sin 2T \cdot \cos(4x+2T) = 0. \text{ Это равенство будет выполнено}$$

$$\text{для любых } x, \text{ если } \sin 2T = 0, \quad 2T = \pi, \quad \boxed{T = \frac{\pi}{2}}.$$

Можно доказать справедливость следующих утверждений:

- Если $k \neq 0$, b – любое число, то для функций

$$\begin{cases} f(x) = \sin(kx+b) \\ f(x) = \cos(kx+b) \end{cases} \quad T = \frac{2\pi}{k}.$$

- Добавление к функции произвольной постоянной и умножение её на произвольную постоянную не меняет периода функции.
- Если функции $y = f_1(x)$ и $y = f_2(x)$ имеют **одинаковый** период, то $y = f_1(x) \pm f_2(x)$ имеет **тот же** период.
- Пусть функции $y = f_1(x)$ и $y = f_2(x)$ имеют соответственно периоды T_1 и T_2 . Если найдётся положительное число T , **кратное T_1 и T_2** , т.е.

$$T = n_1 T_1 = n_2 T_2 \text{ (где } n_1 \text{ и } n_2 \text{ - целые числа), то оно будет периодом функции } y = f_1(x) \pm f_2(x).$$

Пример. Найти период функции $f(x) = \sin 2x + \cos 3x$.

Решение. Периоды функций $f_1(x) = \sin 2x$ и $f_2(x) = \cos 3x$ соответственно равны: $T_1 = \frac{2\pi}{2} = \pi$, $T_2 = \frac{2\pi}{3}$. Найдём наименьшее общее кратное чисел T_1 и T_2 :

$T = n_1 \cdot T_1 = n_2 \cdot T_2$, $n_1 \cdot \pi = n_2 \cdot \frac{2\pi}{3}$, $3n_1 = 2n_2$. Наименьшие значения n_1 и n_2 , удовлетворяющие этому равенству, $n_1 = 2$, $n_2 = 3$. $\boxed{T = n_1 T_1 = 2\pi}$.

Сложная функция

Пусть имеется функция $y = f(u)$, где $y \in Y$, $u \in U$. Переменная $u = \varphi(x)$, где $x \in X$. Функция $y = f(\varphi(x))$ называется сложной функцией. Например, $y = \lg \sin x$ – сложная функция, т.к. её можно представить в виде $y = \lg u$, где $u = \sin x$.

Предел функции

Пусть функция $y = f(x)$ задана в некоторой окрестности точки x_0 , кроме, может быть, самой точки x_0 .

1. Число A называется **пределом** функции $f(x)$ в точке $x = x_0$ (или при $x \rightarrow x_0$), если для любого сколь угодно малого числа $\varepsilon > 0$ найдётся такое зависящее от ε число $\delta > 0$, что для всех $x \neq x_0$ и удовлетворяющих неравенству $|x - x_0| < \delta$ выполняется неравенство $|f(x) - A| < \varepsilon$. Этот предел обозначается $\lim_{x \rightarrow x_0} f(x) = A$.

2. Число A_1 называется **пределом** функции $f(x)$ **слева** в точке $x = x_0$, если для любого числа $\varepsilon > 0$ существует число $\delta > 0$ такое, что при $x \in (x_0 - \delta, x_0)$ выполняется неравенство $|f(x) - A_1| < \varepsilon$. Предел слева записывают так: $\lim_{x \rightarrow x_0 - 0} f(x) = A_1$ либо $f(x_0 - 0) = A_1$.

3. Число A_2 называется **пределом** функции $f(x)$ **справа** в точке $x = x_0$, если для любого числа $\varepsilon > 0$ существует число $\delta > 0$ такое, что при $x \in (x_0, x_0 + \delta)$ выполняется неравенство $|f(x) - A_2| < \varepsilon$. Предел справа записывают так: $\lim_{x \rightarrow x_0 + 0} f(x) = A_2$ либо $f(x_0 + 0) = A_2$.

Из существования $\lim_{x \rightarrow x_0} f(x) = A$ следует, что существуют оба односторонних предела, причём $A_1 = A_2 = A$ и, наоборот, если существуют оба односторонних предела и $A_1 = A_2 = A$, то существует $\lim_{x \rightarrow x_0} f(x) = A$. Если $A_1 \neq A_2$, то

$\lim_{x \rightarrow x_0} f(x)$ не существует.

4. Пусть функция $y = f(x)$ определена на промежутке $(-\infty; \infty)$. Число A называется **пределом** $f(x)$ при $x \rightarrow \infty$, и обозначается $A = \lim_{x \rightarrow \infty} f(x)$, если для любого числа $\varepsilon > 0$ существует число $M > 0$ такое, что при всех x , удовлетворяющих

неравенству $|x| > M$, выполняется неравенство $|f(x) - A| < \varepsilon$. Это определение предполагает неограниченное возрастание x по абсолютной величине.

Если $x \rightarrow +\infty$, то пишут $A = \lim_{x \rightarrow +\infty} f(x)$,

если $x \rightarrow -\infty$, то пишут $A = \lim_{x \rightarrow -\infty} f(x)$.

Производная функции

Производная выступает как интенсивность изменения некоторого экономического объекта (процесса) по времени или относительно другого исследуемого фактора и характеризует скорость изменения функции.

Пусть функция $y = f(x)$ определена на промежутке X . Возьмём точку $x \in X$.

Дадим значению x приращение Δx , тогда функция получит приращение

$$\Delta y = f(x + \Delta x) - f(x). \quad (1)$$

Производной функции $y = f(x)$ называется **предел отношения приращения функции к приращению аргумента, когда приращение аргумента стремится к нулю**. Производную обозначают y' , y'_x , $f'(x)$, $\frac{dy}{dx}$, $\frac{df(x)}{dx}$. Т.о. по определению производной имеем:

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \quad (2)$$

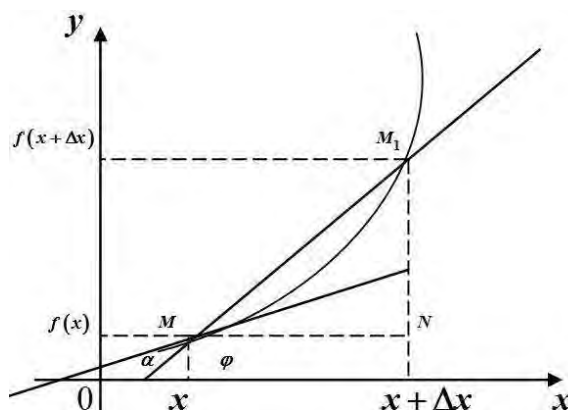
или

$$y' = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}. \quad (3)$$

Функция, имеющая производную в каждой точке интервала, называется **дифференцируемой**, а операция нахождения производной – **дифференцированием**.

Выясним **геометрический** смысл производной. Пусть на плоскости Oxy дана непрерывная кривая $y = f(x)$. Возьмём на кривой две точки: $M(x, f(x))$ и $M_1(x + \Delta x, f(x + \Delta x))$. Прямую MM_1 , проходящую через эти точки, называют **секущей**. **Касательной** к кривой $y = f(x)$ в точке M называется предельное по-

ложение секущей при $\Delta x \rightarrow 0$. Обозначим φ – угол между секущей MM_1 и осью Ox , α – угол между касательной и осью Ox .



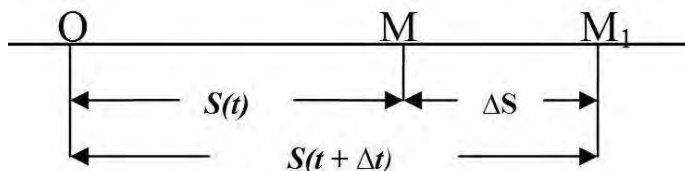
Из $\triangle MM_1N$: $\operatorname{tg} \varphi = \frac{f(x + \Delta x) - f(x)}{\Delta x}$. При $\Delta x \rightarrow 0$ секущая MM_1 переходит в касательную и $\varphi \rightarrow \alpha$ т.е. $\alpha = \lim_{\Delta x \rightarrow 0} \varphi$. Следовательно, $\lim_{\Delta x \rightarrow 0} \operatorname{tg} \varphi = \operatorname{tg} \alpha$ т.е.

$$\operatorname{tg} \alpha = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = f'(x).$$

Геометрический смысл производной: производная $f'(x)$ в точке x равна тангенсу угла наклона касательной к графику функции $y = f(x)$ в этой точке.

Механический смысл производной

Пусть материальная точка M неравномерно движется по некоторой прямой. Каждому значению времени t соответствует расстояние $OM = S$ до некоторой фиксированной точки O .



Это расстояние зависит от времени, т.е. $S = S(t)$. Найдём скорость движения точки. Если в момент времени t точка занимает положение M , то в момент $t + \Delta t$ она займёт положение M_1 , где $OM_1 = S + \Delta S$. Значит перемещение точки за время Δt

будет $\Delta S = S(t + \Delta t) - S(t)$. Отношение $\frac{\Delta S}{\Delta t}$ выражает среднюю скорость движе-

ния точки за время Δt : $v_{cp} = \frac{\Delta S}{\Delta t}$. Предел средней скорости при $\Delta t \rightarrow 0$ называ-

ется скоростью движения точки в данный момент времени или **мгновенной** скоростью. Обозначая её v , получим $v = \lim_{\Delta t \rightarrow 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} = S'(t)$.

Механический смысл производной: скорость прямолинейного движения материальной точки в момент времени t равна производной от пути S по времени t .

В экономике производная равна скорости изменения экономического процесса относительно исследуемого фактора, например, производительность труда в момент времени t равна производной объёма выпущенной продукции по времени.

Связь между непрерывностью и дифференцируемостью функции:

Если функция дифференцируема в некоторой точке, то она непрерывна в этой точке. Обратное утверждение неверно: Непрерывность – необходимое, но недостаточное условие дифференцируемости функции.

Вычисление производных

Производную функции $y = f(x)$ можно найти по следующей схеме:

1. Даём аргументу x приращение $\Delta x \neq 0$ и находим значение функции $f(x + \Delta x)$.
2. Вычисляем $\Delta y = f(x + \Delta x) - f(x)$.
3. Составляем отношение $\frac{\Delta y}{\Delta x}$.
4. Находим предел этого отношения при $\Delta x \rightarrow 0$, т.е. $y' = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}$.

По указанной схеме получена таблица производных. В ней $u = u(x)$, $c = const$.

- | | |
|---|----------------------------|
| 1. $c' = 0$ | 2. $x' = 1$ |
| 3. $(u^\alpha)' = \alpha \cdot u^{\alpha-1} \cdot u'$ | 4. $(e^u)' = e^u \cdot u'$ |

$$5. \quad (a^u)' = a^u \cdot \ln a \cdot u'$$

$$6. \quad (\sin u)' = \cos u \cdot u'$$

$$7. \quad (\cos u)' = -\sin u \cdot u'$$

$$8. \quad (\operatorname{tg} u)' = \frac{1}{\cos^2 u} \cdot u'$$

$$9. \quad (\operatorname{ctg} u)' = -\frac{1}{\sin^2 u} \cdot u'$$

$$10. \quad (\arcsin u)' = \frac{1}{\sqrt{1-u^2}} \cdot u'$$

$$11. \quad (\arccos u)' = -\frac{1}{\sqrt{1-u^2}} \cdot u'$$

$$12. \quad (\operatorname{arctg} u)' = \frac{1}{1+u^2} \cdot u'$$

$$13. \quad (\operatorname{arcctg} u)' = -\frac{1}{1+u^2} \cdot u'$$

$$14. \quad (\ln u)' = \frac{1}{u} \cdot u'$$

$$15. \quad (\log_a u)' = \frac{1}{u \ln a} \cdot u'$$

Правила дифференцирования

Пусть $u = u(x)$, $v = v(x)$, $w = w(x)$, $y = f(u)$, $c = \operatorname{const}$. Тогда

$$1. \quad (u \pm v)' = u' \pm v'$$

$$2. \quad (u \cdot v)' = u'v + uv'$$

$$3. \quad (c \cdot u)' = c \cdot u'$$

$$4. \quad \left(\frac{u}{v} \right)' = \frac{u'v - uv'}{v^2}$$

Если функция $y = f(x)$ задана параметрически, т.е. $x = x(t)$, $y = y(t)$, то $y'_x = \frac{y'_t}{x'_t}$.

Дифференцирование сложной функции: $[f(u)]' = f'_u \cdot u'$ или $y'_x = y'_u \cdot u'_x$.

Эластичность функции

Пусть $y = f(x)$. **Эластичностью** $E_x(y)$ функции y относительно x называется

предел отношения относительного приращения функции $\frac{\Delta y}{y}$ к относительному

приращению аргумента $\frac{\Delta x}{x}$ при $\Delta x \rightarrow 0$:

$$E_x(y) = \lim_{\Delta x \rightarrow 0} \left(\frac{\Delta y}{y} : \frac{\Delta x}{x} \right) = \frac{x}{y} \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \frac{x}{y} \cdot y',$$

$$E_x(y) = \frac{x}{y} \cdot y'.$$

Эластичность приближённо показывает **процентное изменение функции при изменении аргумента на 1%** и используется в экономике при анализе спроса и потребления. Например, эластичность спроса (или количества покупаемого товара) “у” относительно цены товара (или дохода) “х” показывает, на сколько % изменится спрос (или объём потребления) при изменении цены (или дохода) на 1 %.

Спрос называют эластичным при $|E_x(y)| > 1$ и неэластичным при $|E_x(y)| < 1$.

При $|E_x(y)| = 1$ говорят о спросе с единичной эластичностью.

Эластичный спрос означает, что малому процентному изменению цены соответствует большее процентное изменение спроса, т.е. $\frac{\Delta y}{y} > \frac{\Delta x}{x}$, а неэластичный – что малое процентное изменение цены приводит к ещё меньшему процентному изменению спроса.

Пример. Зависимость между себестоимостью единицы продукции “у” (тыс. руб.) и выпуском продукции “х” (млн. руб.) выражается функцией $y = -0.5x + 80$. Найти эластичность себестоимости при выпуске продукции, равном 60 млн. руб.

$$\text{Решение. } E_x(y) = \frac{x}{y} \cdot y' = \frac{x}{-0.5x + 80} \cdot (-0.5x + 80)' = \frac{-0.5x}{-0.5x + 80} = \frac{x}{x - 160}.$$

При $x = 60$, $E_{x=60}(y) = \frac{60}{60 - 160} = -0.6$, т.е. при выпуске продукции, равном 60 млн. руб., увеличение выпуска на 1 % приведёт к снижению себестоимости на 0.6 %.

Непрерывность функции вещественной переменной

Функция $y = f(x)$ называется **непрерывной в точке** x_0 , если она

удовлетворяет трём условиям:

- а) Функция определена в точке x_0 , т.е. существует $f(x_0)$.
- б) Функция имеет конечный предел при $x \rightarrow x_0$.
- с) Этот предел равен значению функции в точке x_0 , т.е. $\lim_{x \rightarrow x_0} f(x) = f(x_0)$.

Данную формулу можно записать в виде $\lim_{x \rightarrow x_0} f(x) = f\left(\lim_{x \rightarrow x_0} x\right)$.

Это означает, что для непрерывной функции **можно менять местами символы предела и функции**.

Если нарушено хотя бы одно из перечисленных условий, то функция не непрерывна. Точки, в которых нарушается непрерывность, называются **точками разрыва функции**.

Можно доказать, что **все элементарные функции непрерывны при всех значениях x , для которых они определены**.

Классификация точек разрыва функции

Пусть существуют в точке $x = x_0$ конечные пределы слева (A_1) и справа (A_2)

функции $y = f(x)$, где $A_1 = \lim_{x \rightarrow x_0 - 0} f(x)$, $A_2 = \lim_{x \rightarrow x_0 + 0} f(x)$. Тогда

1. Если $A_1 = A_2 \neq f(x_0)$, то точка x_0 — точка **устранимого разрыва I рода**.
2. Если $A_1 \neq A_2$, то точка x_0 — точка **неустранимого разрыва I рода**.
3. Если хотя бы один из односторонних пределов не существует или равен ∞ в точке x_0 , то x_0 называется точкой **разрыва II рода**.

Теоремы о непрерывных функциях

1. Сумма, произведение и частное двух непрерывных функций является непрерывной функцией (за исключением тех значений аргумента, при которых знаменатель равен нулю).

2. Если функция $u = \varphi(x)$ непрерывна в точке x_0 , а функция $y = f(u)$ непрерывна в точке $u_0 = \varphi(x_0)$, то сложная функция $f(\varphi(x))$ непрерывна в точке x_0 .
3. Если функция $y = f(x)$ непрерывна и строго монотонна на отрезке $[a, b]$ оси Ox то обратная функция $x = \varphi(y)$ также непрерывна и монотонна на соответствующем отрезке $[c, d]$ оси Oy .

Исследование функции с помощью производных

Достаточные условия возрастания и убывания функции на промежутке:

Функция $y = f(x)$ убывает на промежутке, если внутри него $f'(x) < 0$, и возрастает, если $f'(x) > 0$.

Необходимое условие монотонности функции на промежутке:

Если функция возрастает на некотором промежутке, то $f'(x) \geq 0$ на этом промежутке, если убывает, то $f'(x) \leq 0$, т.е. в отдельных точках может равняться нулю.

Пример. Исследовать функцию $f(x) = x^3 - 3x - 4$ на возрастание и убывание.

Решение. ООФ: $(-\infty, +\infty)$, $f'(x) = 3x^2 - 3 = 3(x-1)(x+1)$, $f'(x) = 0$ при $x = 1$ и $x = -1$. Отметим знаки $f'(x)$ справа и слева от каждой из этих точек – рис.3.



Рис. 3

Ответ: функция возрастает на интервалах $(-\infty, -1)$ и $(1, +\infty)$ и убывает на интервале $(-1, 1)$.

Экстремумы функций

Точка x_0 называется точкой **максимума** функции $f(x)$, если в некоторой её окрестности $f(x) \leq f(x_0)$.

Точка x_1 называется точкой **минимума** функции $f(x)$, некоторой её окрестности $f(x) \geq f(x_1)$.

Значения функции в точках x_0 и x_1 называются соответственно **максимум** и **минимум** функции и объединяются общим названием **экстремум** функции.

Поскольку понятие экстремума связано с достаточно малой окрестностью точки, его часто называют **локальным**. На одном промежутке (a, b) функция может иметь несколько экстремумов – рис. 4.

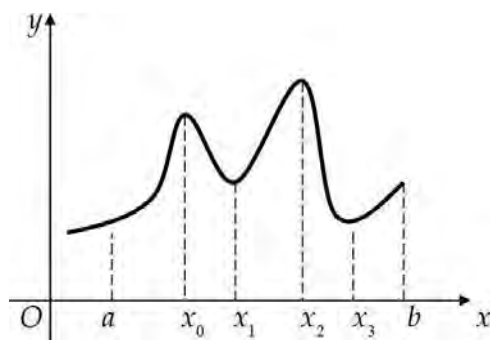


Рис. 4

На рис 4 на промежутке (a, b) имеется четыре локальных экстремума – 2 максимума (в точках x_0 и x_2) и 2 минимума (в точках x_1 и x_3).

Наибольшее и наименьшее значения функции на промежутке называются **глобальным экстремумом**, т.е. наименьший минимум и наибольший максимум среди всех локальных.

1. Необходимое условие экстремума: для наличия экстремума функции $y = f(x)$ в точке x_0 необходимо, чтобы её производная в этой точке **равнялась нулю или не существовала**.

Точки, в которых выполнено необходимое условие экстремума, называются **критическими** или **стационарными**. Они должны входить в область определения функции.

2. Достаточные условия экстремума функции:

- а) Пусть $f(x)$ – дифференцируемая функция. Тогда, если $f'(x)$ при переходе через точку x_0 меняет знак с “+” на “–”, то x_0 – точка локального максимума, а если с “–” на “+”, то x_0 – точка локального минимума.

- б) Пусть $f(x)$ – дважды дифференцируемая функция и $f'(x_0) = 0$. Тогда, если $f''(x_0) > 0$, то x_0 – точка минимума, если $f''(x_0) < 0$, то x_0 – точка локального максимума.

Задача отыскания глобального экстремума функции на отрезке

Нахождение наибольшего и наименьшего значений функции применяется при решении многих практических задач, например, транспортная задача о перевозке груза с наименьшими затратами, организация производственного процесса с целью получения максимальной прибыли и т.д.

Пусть функция $y = f(x)$ непрерывна на отрезке $[a, b]$. Своё наибольшее и наименьшее значение она может принять либо во внутренней точке x_0 отрезка $[a, b]$, либо на границе отрезка, т.е. при $x_0 = a$ или $x_0 = b$. Если $x_0 \in (a, b)$, то x_0 следует искать среди критических точек функции. Отсюда получаем правило нахождения наибольшего и наименьшего значения функции на отрезке $[a, b]$:

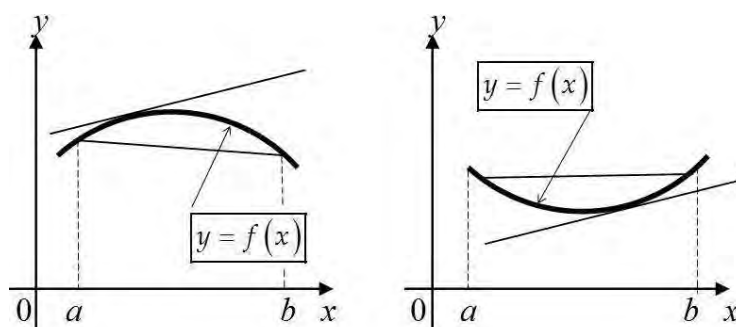
1. Найти критические точки на интервале (a, b) .
2. Вычислить значения функции в критических точках.
3. Вычислить значения функции на концах отрезка $[a, b]$.
4. Среди всех вычисленных значений выбрать наибольшее и наименьшее.

Замечание. Если критических точек на отрезке $[a, b]$ нет, это означает, что функция на этом отрезке монотонна и на одном его конце принимает наибольшее значение, а на другом – наименьшее.

Выпуклость графика функции. Точки перегиба

График дифференцируемой функции $y = f(x)$ называется **выпуклым вниз** на интервале (a, b) , если он расположен выше любой её касательной на этом интервале, и называется **выпуклым вверх** – если расположен ниже любой её касательной на этом интервале. Любая дуга графика выпуклой вверх на интервале (a, b)

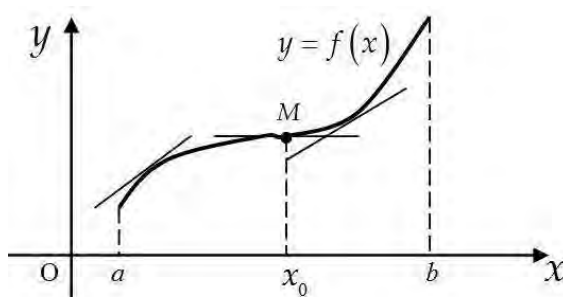
функции лежит выше хорды, соединяющей концы дуги, а дуга графика выпуклой вниз функции – ниже хорды – см. рисунок:



Достаточные условия выпуклости функции $y = f(x)$:

Если $f''(x) > 0$ внутри некоторого промежутка, то функция выпукла вниз на этом промежутке, если $f''(x) < 0$ – то выпукла вверх.

Точкой перегиба называется точка графика, разделяющая интервалы с разной выпуклостью. В окрестности этой точки кривая лежит по разные стороны от касательной, т.е. перегибается через касательную.



На рисунке кривая $y = f(x)$ выпукла вверх на интервале (a, x_0) и выпукла вниз на интервале (x_0, b) , $M(x_0, f(x_0))$ – точка перегиба.

Необходимое условие перегиба в точке $M(x_0, f(x_0))$: $f''(x_0) = 0$.

Достаточное условие перегиба в точке $M(x_0, f(x_0))$: $f''(x)$ меняет знак при переходе через точку x_0 , в которой она равна нулю или не существует.

Асимптоты кривых

Асимптотой графика функции называют прямую, расстояние до которой от

лежащей на кривой точки стремится к нулю при неограниченном удалении этой точки по кривой от начала координат.

1. Вертикальной асимптотой графика функции $y = f(x)$ называют прямую $x = a$, если либо $\lim_{x \rightarrow a} f(x) = \infty$, либо $\lim_{x \rightarrow a+0} f(x) = \infty$, либо $\lim_{x \rightarrow a-0} f(x) = \infty$. Вертикальные асимптоты функции следует искать в **точках разрыва II рода** или на концах её области определения – рис. 1.

2. Горизонтальной асимптотой графика функции $y = f(x)$ называют прямую $y = b$, если существует конечный предел $\lim_{x \rightarrow \infty} f(x) = b$ – рис. 2.

3. Наклонной асимптотой графика функции $y = f(x)$ называют прямую $y = kx + b$, если $f(x)$ определена при достаточно больших x и существуют конечные пределы $k = \lim_{x \rightarrow \infty} \frac{f(x)}{x} \neq 0$, $b = \lim_{x \rightarrow \infty} [f(x) - kx]$ – рис. 3. Если хотя бы один из этих пределов не существует или равен ∞ , то наклонных асимптот нет.

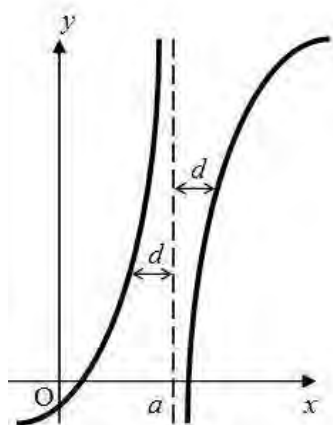


Рис. 1

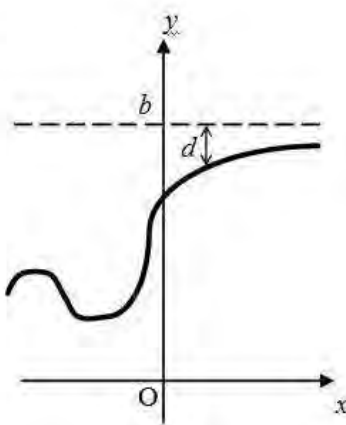


Рис. 2

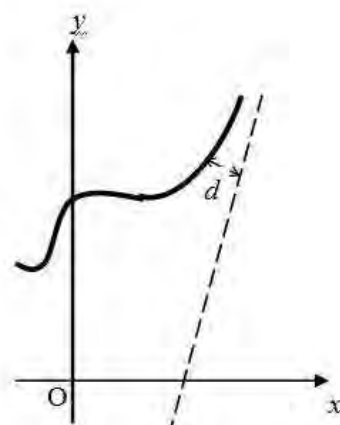


Рис. 3

Замечание. Если конечен только один из пределов при $x \rightarrow +\infty$ или при $x \rightarrow -\infty$, то наклонные асимптоты могут быть односторонними (рис. 2, 3). Горизонтальная асимптота – частный случай наклонной при $k = 0$. Асимптоты могут быть разными при $x \rightarrow +\infty$ и при $x \rightarrow -\infty$, поэтому эти случаи рассматривают отдельно.

Функции многих переменных

Точечные множества в N-мерном пространстве

Множество всевозможных упорядоченных совокупностей n вещественных чисел (x_1, x_2, \dots, x_n) называют **N-мерным координатным пространством**. Каждую такую упорядоченную совокупность называют **точкой n-мерного пространства**, а сами числа – ее **координатами**.

Например, плоскость – двумерное координатное пространство, в котором любая совокупность двух вещественных чисел определяет точку (координаты точки на плоскости можно обозначить (x_1, x_2) , а не только (x, y) , как это было принято при изучении курса планиметрии в рамках школьной программы). Прямая – одномерное координатное пространство. Координаты точки в трехмерном пространстве можно обозначить (x_1, x_2, x_3) или (x, y, z) . Для координат можно использовать различные обозначения, но при этом число координат должно соответствовать размерности пространства (т.е. в двумерном пространстве – две координаты, на прямой – одна координата, в трехмерном пространстве – три координаты, в десятимерном – десять координат и т.д.). Отметим, что если пространства размерности до трех включительно можно зрительно представить себе и даже изобразить, то пространства большей размерности представляют собой научную абстракцию.

N-мерное координатное пространство называют **евклидовым**, если между двумя любыми его точками $X^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)})$ и $X^{(2)} = (x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)})$ определено расстояние, определяющееся соотношением

$$\sqrt{(x_1^{(1)} - x_1^{(2)})^2 + (x_2^{(1)} - x_2^{(2)})^2 + \dots + (x_n^{(1)} - x_n^{(2)})^2}.$$

Множество всех точек n -мерного пространства, удаленных от заданной точки $X^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$ на расстояние, меньшее R , называют **открытым n-мерным шаром** радиуса R с центром в точке $X^{(0)}$, т.е. для всех точек открытого шара $\sqrt{(x_1 - x_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + \dots + (x_n - x_n^{(0)})^2} < R$.

Если это неравенство выполняется, как нестрогое (т.е. расстояние не больше R : $\sqrt{(x_1 - x_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + \dots (x_n - x_n^{(0)})^2} \leq R$), то шар называют **замкнутым**, или просто шаром.

Множество всех точек пространства, равноудаленных от заданной точки, называют **сферой** с центром в заданной точке, т.е. для любой точки сферы $\sqrt{(x_1 - x_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + \dots (x_n - x_n^{(0)})^2} = R$, где R – радиус сферы.

Например, замкнутый шар на плоскости представляет собой круг, т.е. множество точек, удаленных от центра на расстояние, не большее радиуса. Сфера на плоскости представляет собой окружность. Замкнутый шар на прямой – это отрезок (центр – его середина, радиус – половина длины). Сфера – концы этого отрезка. В трехмерном пространстве шар и сферу легко представить себе визуально. В пространствах большей размерности они представляют собой научную абстракцию.

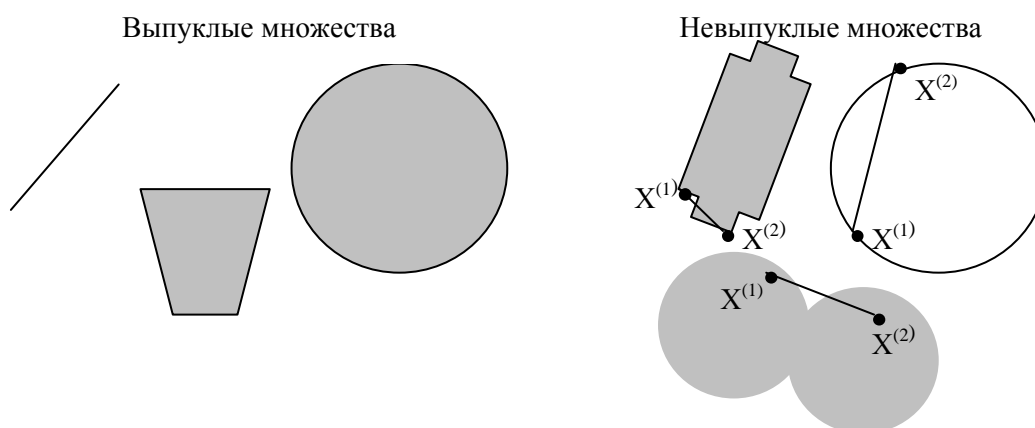
Следует отметить, что если к открытому шару присоединить сферу того же радиуса с тем же центром, то будет получен замкнутый шар. Например, круг на плоскости – это открытый круг вместе с окружностью.

Всякий шар, содержащий точку $X^{(0)}$, называется **окрестностью** точки $X^{(0)}$. Открытый шар радиуса $\varepsilon > 0$ с центром в точке $X^{(0)}$ называют **ε -окрестностью** точки $X^{(0)}$. Точки, в любой ε -окрестности которых содержатся точки, как принадлежащие множеству, так и не принадлежащие ему, называются **граничными**. Например, для шара любая точка соответствующей сферы (с тем же центром и радиусом) является граничной. Если множество содержит все свои граничные точки, оно называется **замкнутым**.

Точка является **внутренней** для некоторого множества, если существует некоторая ее ε -окрестность, все точки которой принадлежат этому множеству. Точка является **внешней** для некоторого множества, если существует некоторая ее окрестность, все точки которой не принадлежат этому множеству. Граничные точки не являются ни внешними, ни внутренними.

Множество точек D n -мерного пространства называется **выпуклым**, если для любых двух точек $X^{(1)}$ и $X^{(2)}$, принадлежащих этому множеству, отрезок, соединяющий эти точки, также, целиком принадлежит этому множеству, т.е. для любых $X^{(1)}, X^{(2)} \in D$ точка $X = \alpha X^{(1)} + (1 - \alpha)X^{(2)} \in D$, где $\alpha \in [0;1]$.

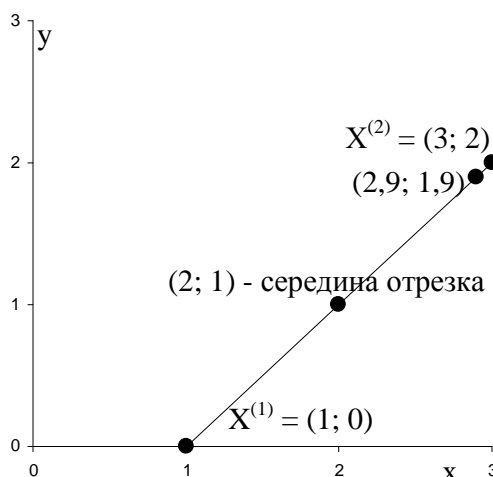
Например, круг или отрезок – выпуклые множества точек, а окружность – невыпуклое. На рисунке ниже изображены примеры фигур, множества точек которых относятся к выпуклым или невыпуклым множествам.



В определении выпуклого множества формула $X = \alpha X^{(1)} + (1 - \alpha)X^{(2)}$ ($\alpha \in [0;1]$) представляет собой формулу отрезка с концами $X^{(1)}$ и $X^{(2)}$, которая при значении параметра $\alpha = 0$ приводит к получению конца $X^{(2)}$, а при значении параметра $\alpha = 1$ приводит к получению конца $X^{(1)}$. При любых других значениях $\alpha \in [0;1]$ будет получена внутренняя точка отрезка, причем если $\alpha = 1/2$, то будет получена его середина, если $\alpha = 1/3$, то отрезок будет разбит этой точкой в пропорции 1:2, начиная от точки $X^{(2)}$ (т.е. будет отсчитана треть длины отрезка от этого конца) и т.д.

Например, возьмем точки $X^{(1)} = (1; 0)$ и $X^{(2)} = (3; 2)$. Формула отрезка между ними примет вид $\alpha X^{(1)} + (1 - \alpha)X^{(2)} = \alpha*(1; 0) + (1 - \alpha)*(3; 2) = (\alpha*1 + (1 - \alpha)*3; \alpha*0 + (1 - \alpha)*2) = (\alpha + 3 - 3\alpha; 2 - 2\alpha) = (3 - 2\alpha; 2 - 2\alpha)$, причем вместо α в этой формуле можно подставлять любое число на промежутке $[0;1]$. Чтобы получить середину отрезка, надо взять $\alpha = 1/2$, в результате чего

мы получим точку $(3 - 2 \cdot 0,5; 2 - 2 \cdot 0,5\alpha) = (2; 1)$. Если взять, например, значение $\alpha = 0,05$, то получим точку $(3 - 2 \cdot 0,05; 2 - 2 \cdot 0,05\alpha) = (2,9; 1,9)$. Она отсчитывает на отрезке $5/100$ или $1/20$ его длины от конца $X^{(2)}$, т.е. разобьет отрезок в пропорции $1:19$. И т.д. Рассмотренные точки отображены на рисунке ниже:



Отметим, что один и тот же отрезок можно описать двумя разными формулами (можно поменять его концы местами, т.е. отсчитывать долю длины отрезка от другого конца).

Понятие функции нескольких переменных

Если каждой точке $X = (x_1, x_2, \dots, x_n)$ из множества $\{X\}$ точек n -мерного пространства ставится в соответствие одно вполне определенное значение переменной величины z , то говорят, что задана **функция n переменных** $z = f(x_1, x_2, \dots, x_n) = f(X)$.

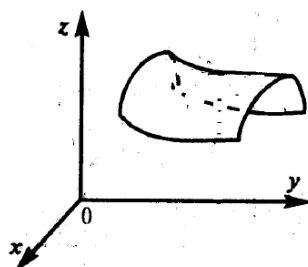
При этом переменные x_1, x_2, \dots, x_n называют **независимыми переменными** или **аргументами** функции, z - **зависимой переменной**, а символ f обозначает **закон соответствия**. Множество $\{X\}$ называют **областью определения** функции (это некое подмножество n -мерного пространства).

Например, функция $z = 1/(x_1 x_2)$ представляет собой функцию двух переменных. Ее аргументы – переменные x_1 и x_2 , а z – зависимая переменная. Область определения – вся координатная плоскость, за исключением прямых

$x_1 = 0$ и $x_2 = 0$, т.е. без осей абсцисс и ординат. Подставив в функцию любую точку из области определения, по закону соответствия получим определенное число. Например, взяв точку $(2; 5)$, т.е. $x_1 = 2$, $x_2 = 5$, получим $z = 1/(2 \cdot 5) = 0,1$ (т.е. $z(2; 5) = 0,1$).

Функция вида $z = a_1x_1 + a_2x_2 + \dots + a_nx_n + b$, где a_1, a_2, \dots, a_n, b - постоянные числа, называют **линейной**. Ее можно рассматривать как сумму n линейных функций от переменных x_1, x_2, \dots, x_n . Все остальные функции называют **нелинейными**. Например, функция $z = 1/(x_1x_2)$ - нелинейная, а функция $z = x_1 + 7x_2 - 5$ - линейная.

Графиком функции двух переменных $z = f(x, y)$ называется множество точек трёхмерного пространства (x, y, z) , аппликата z которых связана с абсциссой x и ординатой y функциональным соотношением $z = f(x, y)$. Этот график представляет собой некоторую поверхность в трехмерном пространстве (например, как на рисунке ниже):



Можно доказать, что если функция - линейная (т.е. $z = ax + by + c$), то ее график представляет собой плоскость в трехмерном пространстве.

Если переменных больше двух (n переменных), то **график** функции представляет собой множество точек $(n+1)$ -мерного пространства, для которых координата x_{n+1} вычисляется в соответствии с заданным функциональным законом. Такой график называют **гиперповерхностью** (для линейной функции - **гиперплоскостью**), и он также представляет собой научную абстракцию (изобразить его невозможно).

Поверхностью уровня функции n переменных называется множество точек в n -мерном пространстве, таких, что во всех этих точках значение функции одно и то же и равно C . Само число C в этом случае называется **уровнем**.

Обычно для одной и той же функции можно построить бесконечно много поверхностей уровня (соответствующих различным уровням).

Для функции двух переменных поверхность уровня принимает вид **линии уровня**.

Например, рассмотрим $z = 1/(x_1x_2)$. Возьмем $C = 10$, т.е. $1/(x_1x_2) = 10$. Тогда $x_2 = 1/(10x_1)$, т.е. на плоскости линия уровня примет вид, представленный на рисунке ниже сплошной линией. Взяв другой уровень, например, $C = 5$, получим линию уровня в виде графика функции $x_2 = 1/(5x_1)$ (на рисунке ниже она показана пунктиром):

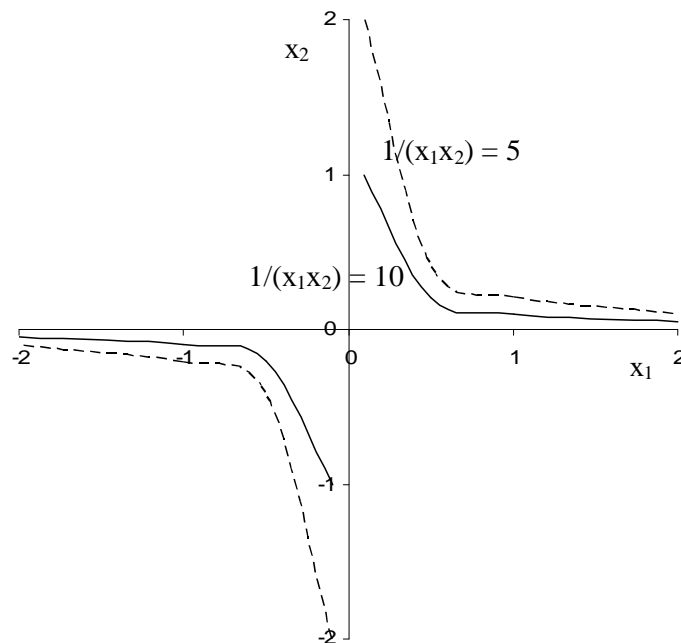


Рисунок - Линии уровня функции $z = 1/(x_1x_2)$

Рассмотрим еще один пример. Пусть $z = 2x_1 + x_2$. Возьмем $C = 2$, т.е. $2x_1 + x_2 = 2$. Тогда $x_2 = 2 - 2x_1$, т.е. на плоскости линия уровня примет вид прямой, представленный на рисунке ниже пунктиром. Взяв другой уровень,

например, $C = 4$, получим линию уровня в виде прямой $x_2 = 4 - 2x_1$ (на рисунке ниже показана сплошной линией). Линия уровня для $2x_1 + x_2 = 3$ показана на рисунке ниже точечной линией.

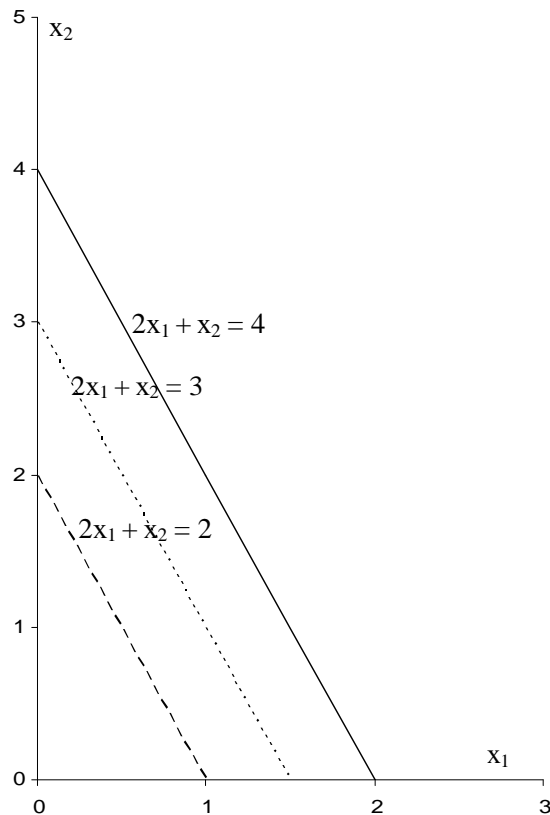


Рисунок - Линии уровня функции $z = 2x_1 + x_2$

Легко убедиться, что для линейной функции двух переменных любая линия уровня будет представлять собой прямую на плоскости, причем все линии уровня будут параллельны между собой.

Частные производные функции многих переменных

Возьмем точку $X = (x_1, x_2, \dots, x_n)$. Дадим аргументу x_1 приращение Δx_1 , аргументу x_2 приращение Δx_2 и т.д., аргументу x_n приращение Δx_n ; тогда функция $z = f(x)$ получит приращение $\Delta z = f(x_1 + \Delta x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n) - f(X)$. Эту величину называются **полным приращением** функции в точке X . Если задать приращение только одного из аргументов, то полученные приращения функции называются **частными**. Например,

$$\Delta_{x_1} z = f(x_1 + \Delta x_1, x_2, \dots, x_n) - f(X),$$

$$\Delta_{x_2} z = f(x_1, x_2 + \Delta x_2, \dots, x_n) - f(X),$$

$$\Delta_{x_n} z = f(x_1, x_2, \dots, x_n + \Delta x_n) - f(X) - \text{частные приращения.}$$

Частной производной функции нескольких переменных $z = f(X)$ называют предел отношения соответствующего частного приращения функции к приращению рассматриваемого аргумента при стремлении последнего к нулю (если этот предел существует):

$$\lim_{\Delta x_j \rightarrow 0} \frac{\Delta_{x_j} z}{\Delta x_j} = \lim_{\Delta x_j \rightarrow 0} \frac{f(x_1, x_2, \dots, x_j + \Delta x_j, \dots, x_n) - f(X)}{\Delta x_j}.$$

Частную производную обозначают z'_{x_j} или $\partial z / \partial x_j$.

Из определения частных производных следует, что для нахождения производной $\partial z / \partial x_j$ надо считать постоянными все переменные аргументы, кроме одного - x_j .

Например, найдем частные производные следующих функций:

Пример 1. $z = x \ln y + y/x$

Чтобы найти частную производную по x , считаем y постоянной величиной. Тогда, $z'_x = \ln y * (x)' + y * (1/x)' = \ln y + y * (-1) * x^{-2} = \ln y - y/(x^2)$.

Аналогично продифференцируем эту функцию по y , считая x постоянной: $z'_y = x (\ln y)' + (1/x) * (y)' = x/y + 1/x$

Пример 2. $z = x^y$

Частная производная по x представляет собой производную степенной функции, т.е. $z'_x = yx^{y-1}$.

Частная производная по y представляет собой производную показательной функции, т.е. $z'_y = x^y \ln x$.

Понятие частной производной имеет вполне четкий экономический смысл. Поскольку функции нескольких переменных в экономике выражают зависимость некоторой величины от нескольких других факторов (иногда

включая время), частная производная выступает как скорость изменения этой величины во времени или относительного другого исследуемого фактора при условии, что остальные факторы не меняются.

Например, пусть магазин продает мороженое – сливочное по 25 руб. за штуку, шоколадное по 30 руб. за штуку и фисташковое по 32 руб. за штуку. Обозначим x_1 – объем продаж сливочного мороженого (шт.), x_2 – объем продаж шоколадного мороженого (шт.), x_3 – объем продаж фисташкового мороженого (шт.). Тогда выручку z (руб.) магазина от продажи этих сортов мороженого можно рассчитать с помощью функции трех переменных $z = 25x_1 + 30x_2 + 32x_3$. Найдем частную производную этой функции по x_1 : $z'_x = 25$. Каков экономический смысл этой величины? Она показывает, на сколько возрастет выручка при единичном изменении продаж сливочного мороженого, при условии, что продажи остальных видов мороженого останутся на прежнем уровне. Иными словами, это скорость изменения общей выручки относительно изменения продаж сливочного мороженого. Аналогичные рассуждения можно провести для обеих других переменных.

Градиент функции

Из школьного курса математики известно, что вектор на плоскости представляет собой направленный отрезок. Его начало и конец имеют по две координаты. Координаты вектора рассчитываются путем вычитания из координат конца координат начала. Понятие вектора может быть распространено и на n -мерное пространство (вместо двух координат будет n координат).

Градиентом $\text{grad } z$ функции $z = f(x_1, x_2, \dots, x_n)$ называется вектор частных производных функции в точке, т.е. вектор с координатами $(z'_{x_1}, z'_{x_2}, \dots, z'_{x_n})$.

Можно доказать, что градиент функции характеризует направление наискорейшего роста уровня функции в точке.

Например, для функции $z = 2x_1 + x_2$ (см. рисунок ниже) градиент в любой точке будет иметь координаты $(2; 1)$. Для линейной функции градиент всегда один и тот же в любой точке. Построить его на плоскости можно различными способами, взяв в качестве начала вектора любую точку. Например, можно соединить точку $(0; 0)$ с точкой $(2; 1)$, или точку $(1; 0)$ с точкой $(3; 1)$, или точку $(0; 3)$ с точкой $(2; 4)$, или т.п. Все построенные таким образом вектора будут иметь координаты $(2 - 0; 1 - 0) = (3 - 1; 1 - 0) = (2 - 0; 4 - 3) = (2; 1)$. Из рисунка хорошо видно, что уровень функции растет в направлении градиента, поскольку построенные линии уровня соответствуют значениям уровня $4 > 3 > 2$.

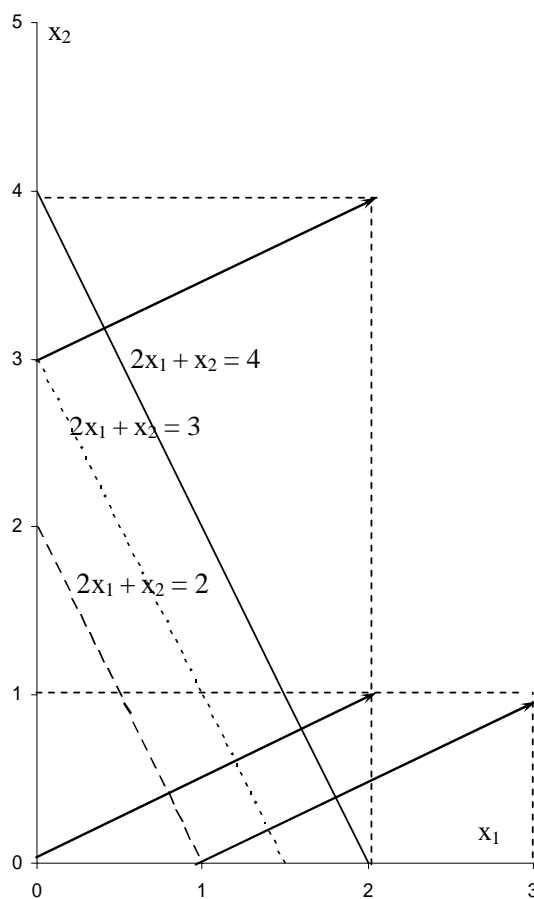


Рисунок - Градиент функции $z = 2x_1 + x_2$

Если взять в качестве примера нелинейную функцию $z = 1/(x_1 x_2)$, то градиент этой функции уже не будет всегда одинаковым в разных точках. Его координаты - не константы, а определяются формулами $(-1/(x_1^2 x_2); -1/(x_1 x_2^2))$.

Явления окружающего нас мира часто протекают под действием множества факторов, среди которых нельзя выделить главные и второстепенные, а проследить влияние всех факторов невозможно. Тогда применяется статистический метод изучения, идея которого заключается в том, что изучается не единичное явление, а массовая совокупность однородных явлений. В такой массовой совокупности влияние каждого второстепенного фактора носит случайный характер и в общей массе взаимно погашается. В результате появляются статистические закономерности. Они не дают возможности предсказать каждое единичное явление, но позволяют достаточно точно описать поведение всей совокупности в целом и на этом основании составить научный прогноз.

С необходимостью изучать массовые совокупности однородных явлений встречаются обычно **в экономике** при анализе темпов роста производства, повышения производительности труда, снижения себестоимости.

Теория вероятностей (ТВ) – математическая наука, изучающая закономерности, присущие массовым случайным явлениям. При этом изучаемые явления рассматриваются в абстрактной форме независимо от их конкретной природы. То есть теория вероятностей рассматривает не сами реальные явления, а их упрощённые схемы – математические модели.

Предметом ТВ являются математические модели случайных явлений, исход которых предсказать невозможно. Например, выпадение герба при подбрасывании монеты, выигрыш по купленному лотерейному билету, результат измерения какой-нибудь величины, длительность работы телевизора и т.п.

Цель ТВ – прогноз в области случайных явлений, влияние на ход этих явлений, контроль их, ограничение сферы действия случайности.

Случайным называется эксперимент, исход которого не вполне однозначно определяется условиями опыта, например, изготовление детали заданных размеров, подбрасывание монеты, вытаскивание карты из колоды.

Пусть A – один из возможных исходов случайного эксперимента, например, выпадение чётного числа очков при подбрасывании игральной кости. Повторяем опыт n раз и пусть при этом исход A наступает m_A раз. Относительной частотой

исхода A наз. величину $W(A) = \frac{m_A}{n}$. Если с увеличением n относительная частота начинает стабилизироваться и при больших n лишь слегка колеблется около некоторого постоянного числа, то говорят, что опыт обладает свойством устойчивости частот.

ТВ изучает математические модели таких случайных экспериментов. Её методы позволяют предсказать средний результат массы случайных экспериментов.

Рассмотрим **основные понятия** теории вероятностей.

Испытание – реальный или принципиально осуществимый опыт, для которого установлены контролируемые воспроизводимые условия и совокупность всех возможных исходов.

Пусть в результате испытания происходит один из множества взаимно исключающих друг друга исходов $\omega_1, \omega_2, \dots, \omega_n$, которые будем называть также элементарными событиями. Они образуют пространство элементарных событий

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}.$$

- **Случайным событием** называют совокупность всех элементарных событий, в результате которых оно наступает. Случайные события обозначают заглавными латинскими буквами A, B, \dots . Говорят, что случайное событие A связано с рассматриваемым испытанием, если по каждому исходу можно точно судить о том, осуществляется оно или нет.

Пример. Испытание – подбрасывание игральной кости. $\Omega = \{1, 2, \dots, 6\}$.

Событие A – выпадение чётного числа очков, $A = \{2, 4, 6\}$,

событие B – выпадение числа очков, кратного трём, $B = \{3, 6\}$.

- Событие \bar{A} называется **противоположным** событию A или дополнением для A , если оно происходит тогда и только тогда, когда не происходит событие A , т.е. содержит все элементарные события, не вошедшие в A . В нашем примере \bar{A} – выпадение нечётного числа очков, $\bar{A} = \{1, 3, 5\}$.
- **Достоверным** называется событие Ω , которое всегда наступает в результате испытания, оно *содержит всю совокупность элементарных событий*,

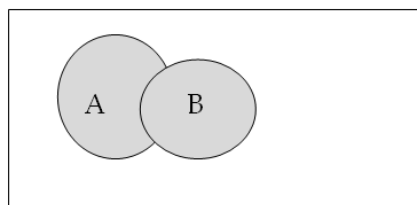
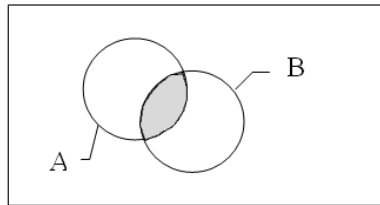
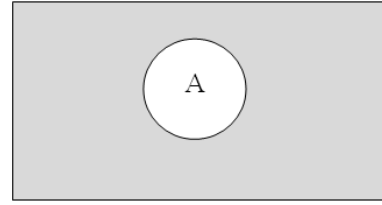
т.е. $U = \Omega$, например, U – выпадение любого числа очков от 1 до 6. Достоверное событие часто обозначают Ω .

- **Невозможным** называется событие V , которое никогда не наступает в результате испытания, оно не содержит ни одного элементарного события, т.е. $V = \{\emptyset\}$, например, V – выпадение 7 очков. Невозможное событие часто обозначают \emptyset .
- Два события называются **несовместными**, если наступление одного из них исключает наступление другого, т.е. **они не содержат общих элементарных событий**, например, при подбрасывании игральной кости $A = \{2, 4, 6\}$ и $C = \{1, 5\}$ несовместны.. Несколько событий называются **попарно несовместными**, если любые два из них несовместны.
- **Суммой** событий A и B называют событие $C = A + B$, состоящее в наступлении хотя бы одного из событий A или B , оно **содержит все элементарные события, входящие в A , в B и общие для A и B** . Например, берём одну карту из колоды, пусть A – появление туза, B – появление красной карты, тогда $C = A + B$ – появление красной карты или любого туза.
- **Произведением** событий A и B называют событие $D = A \cdot B$, состоящее в совместном наступлении A и B , оно **содержит лишь общие для A и B элементарные события**. В предыдущем примере $D = A \cdot B$ – появление красного туза.
- События A_1, A_2, \dots, A_n образуют **полную группу**, если в результате испытания может наступить одно и только одно из них, т.е. у них нет общих исходов (элементарных событий), а в сумме исходы образуют всё пространство элементарных событий:
$$\begin{cases} A_1 + A_2 + \dots + A_n = U \\ A_i \cdot A_j = V, \quad i \neq j \end{cases}.$$

Противоположные события A и \bar{A} всегда образуют полную группу.

При рассмотрении событий удобно использовать **диаграммы**. Пусть точки внутри прямоугольника образуют совокупность всех элементарных событий, полу-

ченных в результате испытания, событие A – попадание в закрашенную область. Тогда сумму событий, произведение и \bar{A} на диаграмме изображают так:


 $A+B$

 $A \cdot B$

 \bar{A}

Операции над событиями обладают следующими свойствами:

1. Переместительное свойство:

$$A+B=B+A, \quad A \cdot B=B \cdot A.$$

2. Сочетательное свойство:

$$(A+B)+C=A+(B+C), \quad (AB)C=A(BC)$$

3. Распределительное свойство:

$$A(B+C)=AB+AC.$$

Справедливы следующие равенства:

$$\begin{aligned} A+\bar{A}=U, & \quad A \cdot \bar{A}=V, & A+A=A, & \quad A \cdot A=A, \\ A+U=U, & \quad A \cdot U=A, & A+V=A, & \quad A \cdot V=V, \\ U+V=U, & \quad U \cdot V=V. \end{aligned}$$

Вероятность и способы её определения

Каждому случайному событию A поставим в соответствие число $P(A)$, называемое вероятностью события A и характеризующее шансы наступления этого события в эксперименте.

• Статистическая вероятность

Рассмотрим эксперимент с пространством элементарных событий Ω и пусть A – случайное событие для этого эксперимента. Повторяем опыт n раз, считая, что результаты каждого не влияют на результаты остальных. Обозначим m_A число

наступлений события A . Дробь $P_n^*(A) = \frac{m_A}{n}$ называется относительной частотой события A в проведённой серии опытов.

- **Статистической вероятностью** события A называется число, около которого колеблется относительная частота события A при **достаточно большом** числе опытов, $P(A) \approx P_n^*(A)$.

Свойства статистической вероятности:

1. $0 \leq P(A) \leq 1$, т.к. $0 \leq m_A \leq n$
2. $P(U) = 1$, т.к. $m_U = n$
3. Для несовместных событий $P(A+B) = P(A) + P(B)$, т.к. $m_{A+B} = m_A + m_B$
4. $P(\bar{A}) = 1 - P(A)$, т.к. $m_{\bar{A}} = n - m_A$.

- **Классическая вероятность**

Рассмотрим эксперимент, у которого n взаимно исключающих друг друга **равновозможных** исходов $\omega_1, \omega_2, \dots, \omega_n$. Классическую вероятность случайного события A определим по формуле

$$P(A) = \frac{m_A}{n},$$

где m_A – количество исходов (элементарных событий), принадлежащих событию A (они называются **благоприятствующими** A). Свойства классической вероятности совпадают со свойствами статистической.

Пример. Определим двумя способами вероятность выпадения чётного числа очков при подбрасывании игральной кости.

1. Многократно подбрасываем игральную кость, например, $n = 500$ и пусть при этом 241 раз выпало чётное число очков, $m_A = 241$. По статистическому определению вероятности $P(A) \approx \frac{241}{500} = 0.482$.

2. Для использования классической вероятности, кость должна быть идеальной, чтобы исходы были равновозможными. Тогда $n = 6$, $m_A = 3$,

$$P(A) = \frac{3}{6} = 0.5.$$

Аксиомы теории вероятностей

В основу ТВ положена система аксиом, опирающаяся на свойства, присущие классическому и статистическому определению вероятности:

1. Каждому событию A ставится в соответствие неотрицательное число $P(A)$, называемое вероятностью этого события.

2. Аксиома сложения: если события A_1, A_2, \dots, A_n попарно несовместны, т.е.

$$A_i \cdot A_j = V \text{ при } i \neq j, \text{ то } P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

3. Аксиома нормировки: если U – достоверное событие, то $P(U) = 1$.

Следствия из аксиом:

a) $P(V) = 0$.

b) $P(\bar{A}) = 1 - P(A)$.

c) Если события A_1, A_2, \dots, A_n образуют полную группу, то $\sum_{i=1}^n P(A_i) = 1$.

d) $0 \leq P(A) \leq 1$.

Вероятность – числовая характеристика степени объективной возможности появления данного события при определённых условиях, которые могут повторяться неограниченное число раз.

Зависимые и независимые события. Условная вероятность

Пусть A и B – события, рассматриваемые в одном опыте. Наступление одного из них может влиять на возможность наступления другого. Вероятность события A при условии, что произошло B , называется **условной** и обозначается $P(A|B)$.

Событие A называется **независимым** от B , если его вероятность не меняется в зависимости от наступления B , т.е. $P(A|B) = P(A)$.

Следует отметить, что **зависимость и независимость событий всегда взаимна**: если A не зависит от B , то и B не зависит от A .

Теорема умножения вероятностей: вероятность совместного наступления двух событий равна произведению вероятности одного из них на условную вероятность второго, вычисленную при условии, что первое событие наступило:

$$P(AB) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B). \quad (1)$$

Из формулы (1) получаем $P(A|B) = \frac{P(AB)}{P(B)}$, (2)

если $P(B) \neq 0$. Формула (2) при аксиоматическом построении теории принимается в качестве определения условной вероятности события A при условии, что произошло событие B . В случае нескольких событий теорема умножения вероятностей примет вид

$$P(A_1 A_2 \dots A_n) = P(A_1) \cdot P(A_2|A_1) \cdot \dots \cdot P(A_n|A_1 A_2 \dots A_{n-1}) \quad (3)$$

Пример. В ящике 2 белых и 3 чёрных шара. Наудачу берут 2. Какова вероятность, что оба белые (событие B)?

Решение. Первый способ: используем теорему умножения вероятностей.

Обозначим A_i – появление белого шара при i – м извлечении,

$$i = 1, 2; \quad B = A_1 \cdot A_2, \quad P(B) = P(A_1)P(A_2|A_1) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}.$$

Второй способ – по классической вероятности: $P(B) = \frac{m_B}{n} = \frac{C_2^2}{C_5^2} = \frac{1}{10}$.

Несколько событий называются **независимыми**, если каждое из них не зависит от произведения любого числа остальных и от каждого в отдельности.

Для независимых событий условные вероятности совпадают с безусловными и

$$P(A_1 A_2 \dots A_n) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n). \quad (4)$$

Пример. Три студента независимо друг от друга решают одну задачу. Вероятности решения задачи каждым из них соответственно равны 0.5, 0.6 и 0.7. Найти вероятность того, что задачу решили только первый и третий студенты (событие А).

Решение. Обозначим B_i ($i = 1, 2, 3$) – задачу решил студент № i. Событие А заключается в совместном наступлении независимых событий $B_1, \overline{B_2}, B_3$. По формуле (4) получим:

$$A = B_1 \cdot \overline{B_2} \cdot B_3, \quad P(A) = P(B_1) \cdot P(\overline{B_2}) \cdot P(B_3) = 0.5 \cdot (1 - 0.6) \cdot 0.7 = 0.14.$$

Теорема сложения вероятностей для произвольных событий:

$$P(A + B) = P(A) + P(B) - P(AB) \quad (5)$$

Пример. Два стрелка с вероятностями попадания соответственно 0.7 и 0.8 стреляют по одному разу в общую мишень. Найти вероятность поражения мишени (событие С).

Решение. Обозначим: событие А – попал первый стрелок, событие В – попал второй. Решим задачу тремя способами.

1. $C = A + B$, причём А и В совместны и независимы. Применяя формулы (5) и (4), получим:

$$P(C) = P(A) + P(B) - P(AB) = P(A) + P(B) - P(A)P(B) = 0.7 + 0.8 - 0.7 \cdot 0.8 = 0.94.$$

2. $C = A\overline{B} + \overline{A}B + AB$, причём слагаемые в правой части равенства являются попарно несовместными событиями, а сомножители – независимые события.

Используя аксиому сложения и формулу (4), получим:

$$P(C) = P(A\bar{B}) + P(\bar{A}B) + P(AB) = P(A)P(\bar{B}) + P(\bar{A})P(B) + P(A)P(B) = \\ = 0.7 \cdot (1 - 0.8) + (1 - 0.7) \cdot 0.8 + 0.7 \cdot 0.8 = 0.14 + 0.24 + 0.56 = 0.94.$$

3. Рассмотрим событие \bar{C} – оба стрелка промахнулись.

$$\bar{C} = \bar{A}\bar{B}, \quad P(\bar{C}) = P(\bar{A})P(\bar{B}) = (1 - 0.7)(1 - 0.8) = 0.3 \cdot 0.2 = 0.06, \\ P(C) = 1 - P(\bar{C}) = 0.94.$$

Замечание: если интересующее нас событие A распадается на много вариантов, то бывает **проще найти вероятность противоположного события \bar{A}** и воспользоваться равенством $P(A) = 1 - P(\bar{A})$, как поступили при решении задачи третьим способом.

Формула полной вероятности

Пусть об условиях наступления события A можно сделать n взаимно исключающих предположений (гипотез) H_1, H_2, \dots, H_n (т.е. A наступает в совокупности с одной и только одной из гипотез). События H_1, H_2, \dots, H_n образуют полную группу, следовательно, $\sum_{i=1}^n P(H_i) = 1$. Тогда вероятность события A вычисляют по **формуле полной вероятности**

$$P(A) = \sum_{i=1}^n P(H_i) \cdot P(A|H_i) \quad (6)$$

При применении этой формулы можно условно разбить опыт на 2 этапа: гипотезы H_1, H_2, \dots, H_n исчерпывают все возможные предположения относительно исходов первого этапа опыта, событие A – один из возможных исходов второго этапа.

Формула Байеса

Пусть об условиях наступления события A можно сделать n взаимно исключающих предположений (гипотез) H_1, H_2, \dots, H_n , и известно, что в результате опыта событие A произошло. Вероятность того, что при этом имела место гипотеза H_k ,

т.е. $P(H_k|A)$, где $k = 1, 2, \dots, n$, вычисляют по **формуле Байеса**:

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{\sum_{i=1}^n P(H_i)P(A|H_i)} = \frac{P(H_k)P(A|H_k)}{P(A)}. \quad (7)$$

Пример. Заказчику поставляют детали 3 фирмы. На долю первой приходится 50% общего объема поставок, а на две другие – соответственно 30% и 20%. Процент брака соответственно равен 10%, 5% и 6%. Из поставленных деталей наугад берется одна деталь. Найти: 1) вероятность того, что эта деталь бракованная, 2) вероятность того, что деталь поставлена первой фирмой, если она бракованная.

Решение. Обозначим: событие A – деталь бракованная. Об условиях наступления события A можно сделать три взаимно исключающих предположения (гипотезы)

H_i – деталь изготовлена i -ой фирмой, где $i = 1, 2, 3$, причём по условию задачи:

$P(H_1) = 0.5$, $P(H_2) = 0.3$, $P(H_3) = 0.2$. Таким образом, опыт проходит в два эта-

па. На первом этапе фирмы поставляют товар в указанных объёмах, а на втором этапе берётся 1 деталь.

Соответствующие условные вероятности того, что она бракованная равны:

$P(A|H_1) = 0.1$, $P(A|H_2) = 0.05$, $P(A|H_3) = 0.06$. По формуле полной вероятности (6) получим:

$P(A) = \sum_{i=1}^3 P(H_i)P(A|H_i) = 0.5 \cdot 0.1 + 0.3 \cdot 0.05 + 0.2 \cdot 0.06 = 0.077$.

Ответ на второй вопрос: $P(H_1|A) = \frac{P(H_1)P(A|H_1)}{P(A)} = \frac{0.5 \cdot 0.1}{0.077} \approx 0.65$.

Схема испытаний Бернулли. Формула Бернулли

Несколько опытов называются **независимыми**, если вероятность того или иного исхода каждого из них не зависит от результатов остальных опытов.

Примеры: 1) несколько последовательных подбрасываний монеты, 2) несколько последовательных извлечений двух шаров из ящика с белыми и чёрными шарами, если извлечённые шары перед каждым новым извлечением снова возвращают в ящик и шары перемешивают.

Рассмотрим серию из n повторных независимых опытов, в каждом из которых может наступить либо некоторое событие A (успех) с вероятностью p , либо \bar{A}

(неудача) с вероятностью $q = 1 - p$. Такие повторные независимые опыты называются **испытаниями Бернулли** или **схемой Бернулли**.

Вероятность $P_{n,k}$ того, что в серии из n испытаний Бернулли событие A наступит ровно k раз, находят по **формуле Бернулли**:

$$P_{n,k} = C_n^k p^k q^{n-k}. \quad (8)$$

Число k_0 , при котором достигается максимальное значение этой вероятности, называется **наивероятнейшим числом наступлений события A** и определяется неравенствами (если $p \neq 0, p \neq 1$): $p(n+1) - 1 \leq k_0 \leq p(n+1)$.

Если $p(n+1)$ – целое число, то наивероятнейших значений два:

$$k_0 = p(n+1), \quad k_0^* = k_0 - 1.$$

Распределение вероятностей $P_{n,k}$ между возможными значениями k от 0 до n называется **биномиальным**, поскольку по формуле бинома Ньютона

$$(q + p)^n = \sum_{k=0}^n C_n^k p^k q^{n-k}. \text{ Учитывая, что } q + p = 1, \text{ получим } \sum_{k=0}^n C_n^k p^k q^{n-k} = 1.$$

Пример. В магазине проходит акция по продаже некоторого изделия. Для получения приза нужно собрать 5 изделий с особым знаком на этикетке. Этикетки с этим знаком имеют 5% изделий. Найти вероятность того, что придется купить 10 изделий (событие В).

Решение. Из постановки задачи следует, что десятое изделие должно иметь особый знак. Следовательно, из предыдущих 9 этот знак имели 4 изделия. Найдем вероятность этого события по формуле Бернулли (8). Сформулируем условие задачи в терминах схемы Бернулли. Проводится серия из 9 повторных независимых опытов (покупка изделия), в каждом из которых событие А (появление изделия с особым знаком) может произойти с одинаковой вероятностью

$$p = \frac{5}{100} = \frac{1}{20}, \quad q = 1 - p = \frac{19}{20}. \text{ Вероятность того, что А наступит 4 раза в 9 опытах}$$

$P_{9,4} = C_9^4 \cdot (1/20)^4 \cdot (19/20)^5 = 0.0006092$. Интересующее нас событие В произойдет, если десятое изделие будет иметь особый знак на этикетке, т.е. наступят сра-

зу 2 независимых события: из первых 9 изделий 4 имели особый знак и 10-ое тоже его имеет. По теореме умножения вероятностей находим:

$$P(B) = p \cdot P_{9,4} = (1/20) \cdot 0.0006092 \approx 0.00003.$$

Случайные величины. Дискретная случайная величина

Большинство явлений и процессов характеризуется количественными параметрами, которые изменяются случайным образом, поэтому случайные величины являются основными объектами изучения и управления.

Под случайной величиной (сокращённо **с.в.**) понимают величину, принимающую в результате опыта то или иное значение, причём заранее неизвестно, какое именно. Случайные величины обозначают заглавными латинскими буквами X, Y, Z, \dots , а принимаемые ими значения – соответствующими малыми буквами x, y, z, \dots

С.в. называется **дискретной** (сокращённо ДСВ), если множество её значений конечное или счётное, она принимает отдельные изолированные друг от друга значения.

Счётным называется бесконечное множество, элементы которого можно перенумеровать с помощью чисел натурального ряда, например, $\left\{1, \frac{1}{4}, \frac{1}{9}, \dots, \frac{1}{n^2}, \dots\right\}$.

Примеры ДСВ: X – число очков, выпадающее при бросании игральной кости, Y – количество выстрелов до первого попадания в цель, Z – число вызовов, поступающих на телефонную станцию в единицу времени, S – число выпадений герба при 10 подбрасываниях монеты.

С.в. называется **непрерывной** (сокращённо НСВ), если множество её значений несчётное, она может принимать любые значения из некоторого промежутка.

Примеры НСВ: X – время безотказной работы прибора, Y – расстояние от центра мишени до пробойны при попадании, Z – дальность полёта снаряда и т.д. Для полного описания случайной величины нужно указать закон её распределения.

Законом распределения случайной величины называется любой способ (табличный, графический или аналитический) задания соответствия между множеством её возможных значений и вероятностями того, что с.в. примет то или иное значение из этого множества.

Универсальным способом задания закона распределения для **любых** случайных величин является задание функции распределения $F(x)$. Для дискретных случайных величин можно также задавать закон распределения в виде ряда распределения или многоугольника распределения, а для непрерывных – в виде плотности вероятности $f(x)$. Рассмотрим подробнее эти способы.

Дискретная случайная величина (ДСВ)

Пусть X – дискретная случайная величина, имеющая n различных возможных значений x_1, x_2, \dots, x_n . При этом случайные события

$$A_1 = \{X = x_1\}, \quad A_2 = \{X = x_2\}, \quad \dots, \quad A_n = \{X = x_n\}$$

образуют полную группу попарно несовместных событий. Обозначим p_1, p_2, \dots, p_n вероятности этих событий, т.е. $p_k = P\{X = x_k\}$, $k = 1, 2, \dots, n$.

Таким образом, каждому возможному значению x_k ДСВ X ставится в соответствие число p_k , т.е. имеется функциональная зависимость между возможными значениями ДСВ и соответствующими вероятностями. Эта зависимость является простейшей формой закона распределения ДСВ.

Таблица, в которой перечислены все возможные значения x_k ДСВ X и указаны их вероятности p_k называется **рядом распределения дискретной случайной величины X** . Ряд распределения изображен на рис. 1.

x_k	x_1	x_2	\dots	x_n
p_k	p_1	p_2	\dots	p_n

Рис. 1

Значения x_k обычно располагают в порядке возрастания. Поскольку события A_1, A_2, \dots, A_n образуют полную группу, сумма их вероятностей равна единице:

$$\sum_{k=1}^n p_k = 1.$$

Графически ряд распределения представляют в виде **многоугольника распределения**, который также называют **полигоном распределения**. При этом

на оси Ox откладывают x_k , на оси Oy – p_k , точки (x_k, p_k) соединяют ломаной.

Пример. В корзине 3 зрелых и 5 незрелых яблок, наудачу берут 3. Написать ряд распределения ДСВ X , равной количеству незрелых яблок среди отобранных.

Решение. ДСВ X может принимать значения 0, 1, 2, 3. Находим соответствующие вероятности:

$$p_1 = P\{X=0\} = \frac{C_5^0 \cdot C_3^3}{C_8^3} = \frac{1}{56}, \quad p_2 = P\{X=1\} = \frac{C_5^1 \cdot C_3^2}{C_8^3} = \frac{15}{56},$$

$$p_3 = P\{X=2\} = \frac{C_5^2 \cdot C_3^1}{C_8^3} = \frac{30}{56}, \quad p_4 = P\{X=3\} = \frac{C_5^3 \cdot C_3^0}{C_8^3} = \frac{10}{56}.$$

Проверка: $\sum_{k=1}^4 p_k = p_1 + p_2 + p_3 + p_4 = 1.$

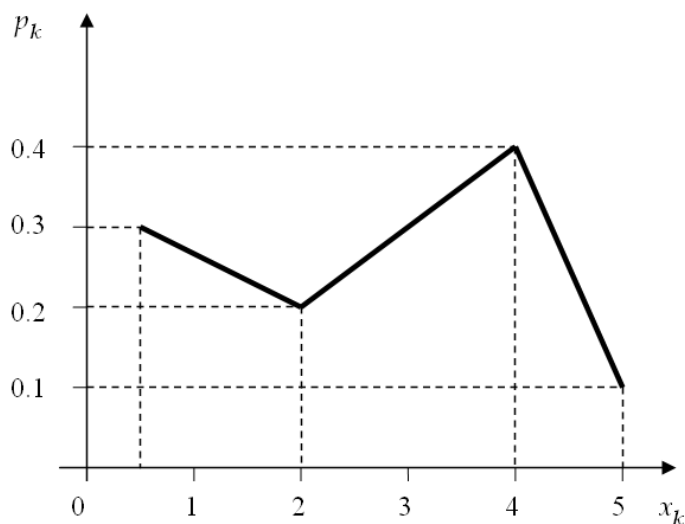
Ряд распределения ДСВ X :

x_k	0	1	2	3
p_k	$\frac{1}{56}$	$\frac{15}{56}$	$\frac{30}{56}$	$\frac{10}{56}$

Пример. Построить многоугольник распределения с.в. X для ряда распределения, заданного таблицей:

x_k	0.5	2	4	5
p_k	0.3	0.2	0.4	0.1

Решение. Многоугольник распределения изображён на рисунке :



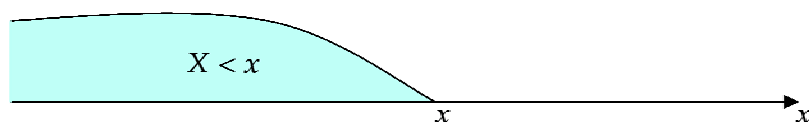
Функция распределения случайной величины и её свойства

Ряд распределения можно построить только для дискретной случайной величины, для непрерывной нельзя даже перечислить все её возможные значения. Универсальным способом задания закона распределения для **любых** случайных величин служит функция распределения.

Функцией распределения $F(x)$ случайной величины X называется функция, которая для любого действительного значения x равна вероятности события $\{X < x\}$, т.е.

$$F(x) = P(X < x), \quad x \in (-\infty, +\infty). \quad (1)$$

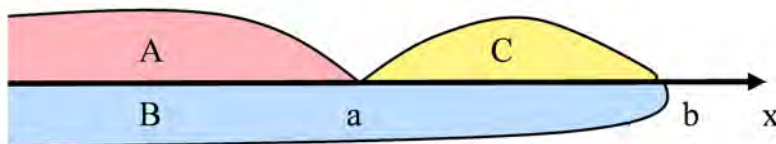
Геометрически равенство (1) можно истолковать так: $F(x)$ – это вероятность того, что с.в. X примет значение, которое на числовой оси изображается точкой, лежащей левее точки x , т.е. случайная точка X попадёт в интервал $(-\infty, x)$:



Свойства функции распределения

1. $0 \leq F(x) \leq 1$, т.к. это вероятность.
2. $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$, т.к. $F(-\infty) = P(X < -\infty) = P(V) = 0$, где V – невозможное событие.
3. $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$, т.к. $F(+\infty) = P(X < +\infty) = P(U) = 1$, где U – достоверное событие.
4. $P(a \leq X < b) = F(b) - F(a)$. Доказательство. Обозначим, как показано на рисунке, события:

$$A = \{X < a\}, \quad B = \{X < b\}, \quad C = \{a \leq X < b\}.$$



Очевидно, $B = A + C$, причём события A и C несовместны. По аксиоме сложения:

$$P(B) = P(A) + P(C). \text{ Получили: } \underbrace{P(X < b)}_{F(b)} = \underbrace{P(X < a)}_{F(a)} + P(a \leq X < b).$$

откуда $P(a \leq X < b) = F(b) - F(a)$ – ч.т.д.

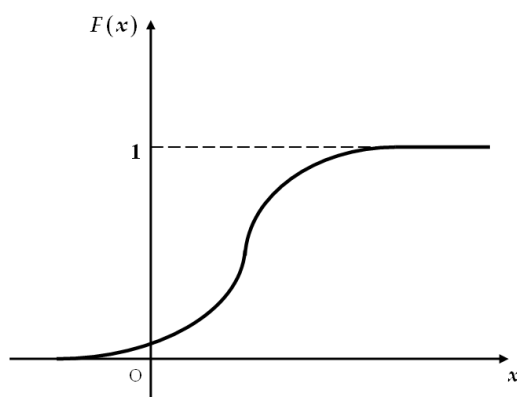
5. $F(x)$ – неубывающая функция, т.е., если $x_1 < x_2$, то $F(x_1) \leq F(x_2)$.

Доказательство. Пусть $x_1 < x_2$. В соответствии с пунктом (4):

$$F(x_2) - F(x_1) = P(x_1 \leq X < x_2) \geq 0,$$

т.к. вероятность не бывает отрицательной. Поэтому $F(x_1) \leq F(x_2)$.

Функция распределения непрерывной случайной величины непрерывна, её график имеет вид:



Функция распределения дискретной случайной величины

Вычислим функцию распределения ДСВ X , имеющей ряд распределения

x_k	x_1	x_2	\dots	x_n
p_k	p_1	p_2	\dots	p_n

Будем присваивать x различные значения и находить для них функцию распределения по формуле: $F(x) = P(X < x)$. Пусть $x_1 < x_2 < \dots < x_n$.

1. Если $x \leq x_1$, то событие $\{X < x\}$ невозможное, т.к. у ДСВ нет значений, меньших x , следовательно, $F(x) = 0$.

2. Если $x_1 < x \leq x_2$, то событие

$$\{X < x\} = \{X = x_1\}, \quad F(x) = P(X < x) = P(X = x_1) = p_1.$$

3. Если $x_2 < x \leq x_3$, то событие $\{X < x\} = \{X = x_1\} + \{X = x_2\}$, причём события в правой части равенства несовместны. По аксиоме сложения:

$$F(x) = P(X < x) = P(X = x_1) + P(X = x_2) = p_1 + p_2.$$

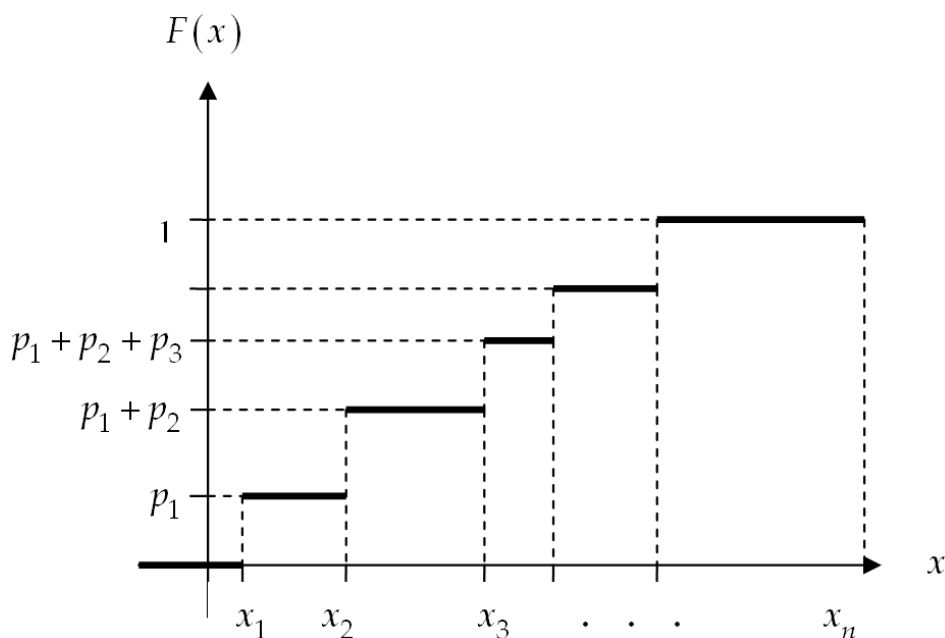
4. Действуя аналогично для каждого следующего промежутка, наконец, получим, что если $x > x_n$, то событие

$$\{X < x\} = \{X = x_1\} + \{X = x_2\} + \dots + \{X = x_n\}, \text{ причём события в правой части равенства несовместны. } F(x) = P(X < x) = p_1 + p_2 + \dots + p_n = 1.$$

Окончательно получили:

$$F(x) = \begin{cases} 0, & x \leq x_1; \\ p_1, & x_1 < x \leq x_2; \\ p_1 + p_2, & x_2 < x \leq x_3; \\ p_1 + p_2 + p_3, & x_3 < x \leq x_4; \\ \dots & \dots \\ 1, & x > x_n. \end{cases} \quad (2)$$

График этой функции имеет вид:



Пример. Найти функцию распределения для ДСВ X , заданной рядом распределения:

x_k	-2.4	3.1	5.2
p_k	0.3	0.1	0.6

Решение. По формуле (2) находим:

$$F(x) = \begin{cases} 0, & x \leq -2.4, \\ 0.3, & -2.4 < x \leq 3.1, \\ 0.4, & 3.1 < x \leq 5.2, \\ 1, & x > 5.2. \end{cases}$$

.

Лекция 12 Статистические показатели и система показателей. Абсолютные и относительные величины

Цель: изучить сущность и значение.

Задачи: изучить классификацию статистических показателей.

План:

1. Функции статистических показателей
2. Механизм расчета.

Любое статистическое исследование в конечном итоге заканчивается расчетом и анализом статистических показателей. Статистический показатель представляет собой количественную характеристику явлений и процессов.

Как правило, изучаемые процессы и явления сложны, их сущность нельзя отразить одним показателем. Поэтому возникает необходимость в применении системы статистических показателей.

Статистический показатель представляет собой количественную характеристику социально-экономических явлений и процессов (признаков).

Система статистических показателей - это совокупность взаимосвязанных между собой показателей, характеризующих социально-экономические явления и направленная на решение конкретной статистической задачи.

В основу классификации статистических показателей положены следующие **признаки**:

1) по охвату единиц совокупности:

- индивидуальные показатели - характеризуют отдельный объект или отдельную единицу совокупности;
- сводные - характеризуют группу единиц, которая представляет собой часть совокупности или всю совокупность в целом.

2) по форме выражения показатели бывают:

- абсолютные;
- относительные;

- средние.

3) по временному фактору:

- моментные;
- интервальные.

4) по принадлежности объекта:

- однообъектные;
- межобъектные.

5) по признаку пространственной определённости:

- общетерриториальные;
- региональные;
- местные.

6) по выполняемой функции:

- плановые показатели – ориентированы на выполнение поставленных задач;
- учетные показатели – показывают реальное состояние изучаемого явления;
- прогностические показатели – показывают возможное состояние в будущем.

Абсолютные статистические показатели – величины, которые в количественном выражении характеризуют какой-либо статистический показатель.

Абсолютные показатели характеризуют численность совокупности, либо объём изучаемого явления в конкретных границах пространства и времени. Они отражают уровень развития явления и его размер.

Абсолютный показатель можно получить одним из двух способов:

- путём подсчёта единиц совокупности, обладающих конкретным значением признака.
- путём суммирования значений признака по всей статистической совокупности.

Абсолютные показатели всегда имеют определённую единицу измерения и размерность.

В зависимости от сущности исследуемых явлений они могут выражаться в натуральных, трудовых и стоимостных единицах измерения. В соответствии с этим различают натуральные, трудовые и стоимостные абсолютные показатели.

Натуральные единицы измерения используются в тех случаях, когда единицы измерения соответствуют потребительским свойствам изучаемых явлений. Например, производство автомобилей в штуках, производство стали в тоннах, урожайность в центнерах и т.д.

Натуральные единицы могут быть составными (сложными). Такие единицы применяются в тех случаях, когда для характеристики изучаемого явления одной единицы недостаточно, и используется произведение двух единиц. Например, производство электроэнергии измеряется в киловатт-часах, грузооборот – в тонно-километрах и т.д.

Стоимостные измерители позволяют дать денежную оценку изучаемым явлениям и процессам.

Эти измерители используются при обобщении данных при оценке неоднородных статистических совокупностей (начиная с уровня предприятия до уровня народного хозяйства). Например, объем выпущенной продукции предприятия, доходы населения и т.д.

Показатели, выраженные в стоимостных единицах измерения можно суммировать, получать по ним итоговые данные, но при их использовании необходимо учитывать изменение цен с течением времени.

Трудовые единицы измерения применяются при оценке общих затрат труда и трудоёмкости операции технологического процесса. К ним относятся человеко-дни, человеко-часы (при оценке затрат рабочего времени) и нормо-минуты (при оценке трудоемкости).

Абсолютные величины отражают наличие тех или иных ресурсов и являются основой материального учёта, однако они не дают полного представления об изучаемом явлении, не показывают его структуру, развитие во времени и соотношение между отдельными частями изучаемой совокупности.

Абсолютные показатели всегда именованные числа, т. е. имеют единицу измерения.

Натуральные единицы измерения применяют в тех случаях, когда единицы измерения соответствуют потребительским свойствам продукта (т, м, шт., мили, унции, галлоны). Натуральные единицы могут быть и составными (сложными). Для того чтобы полнее охарактеризовать потребительское назначение продукции

Если некоторые разновидности продукции обладают общими потребительскими свойствами, обобщенные итоги выражают в условно натуральных единицах (базовая жирность молока, содержание питательного вещества, сопоставимые цены (инфляция)).

Наиболее широко используются стоимостные (денежные) единицы измерения. Для получения общего объема продукции в денежном выражении количество единиц в натуральном выражении умножается на цену, а затем полученную величину суммируют.

Таким образом, абсолютные величины получают непосредственным подсчетом данных статистического наблюдения или расчетным путем.

Абсолютные статистические показатели могут быть измерены с различной степенью точности. Пример: шт., млн. шт.; т, тыс. т, млн. т.

Соблюдение одинаковых единиц измерения – непереносимое условие при сравнениях.

На основе абсолютных показателей сложно проводить сопоставление с другими подобными явлениями. Перечисленные функции выполняют относительные показатели.

Относительные показатели – результат соотношения двух абсолютных показателей. Поэтому, по отношению к абсолютным показателям, относительные показатели являются вторичными.

При расчете относительного показателя, абсолютный показатель (числитель) называется текущим или сравниваемым. Показатель, с которым сравнивают (знаменатель) – основание или база сравнения.

Все относительные статистические показатели классифицируются следующим образом:

- Динамики
- Плана
- Реализации плана
- Структуры
- Координации
- Интенсивности и уровня экономического развития

Сравнения (наглядности)

Относительные показатели динамики (ОПД) – отношение уровня исследуемого процесса за период времени к уровню того же процесса в прошлом.

$$\text{ОПД} = \frac{\text{Текущий показатель}}{\text{Предшествующий или базисный показатель}}$$

Относительный показатель плана (ОПП) – применяется при перспективных расчетах, т.е. планировании.

$$\text{ОПП} = \frac{\text{Показатель, планируемый на (i+1) период}}{\text{Показатель, достигнутый в этом периоде}}$$

При сравнении реально достигнутого результата с ранее намеченным, определяют относительный показатель реализации плана (ОПРП).

$$\text{ОПРП} = \frac{\text{Показатель, достигнутый в (i+1) периоде}}{\text{Показатель, планируемый на (i+1) период}}$$

Между относительным показателем плана (ОПП), реализации плана (ОПРП) и динамики (ОПД) существует следующая взаимосвязь:

$$\text{ОПП} \times \text{ОПРП} = \text{ОПД}$$

Относительный показатель структуры (ОПС) – соотношение структурных частей изучаемого объекта и их целого.

$$\text{ОПС} = \frac{\text{Показатель, характеризующий часть совокупности}}{\text{Показатель по всей совокупности в целом}}$$

Относительный показатель координации (ОПК) – характеризует соотношение отдельных частей целого между собой.

$$\text{ОПК} = \frac{\text{Показатель, характеризующий } i \text{ часть совокупности}}{\text{Показатель, характеризующий часть совокупности, выбранной в качестве базы}}$$

Относительный показатель интенсивности (ОПИ) – характеризует степень распространения изучаемого процесса в присущей ему среде.

$$\text{ОПИ} = \frac{\text{Показатель, характеризующий явление А}}{\text{Показатель, характеризующий среду распространения явления А}}$$

Разновидностью относительного показателя интенсивности является относительные показатели уровня экономического развития, характеризующие производство продукции в расчете на душу населения и играющие важную роль в оценке развития экономики государства.

Относительный показатель сравнения (ОПСр) – соотношение одноименных абсолютных показателей, характеризующих разные объекты (фирмы, районы, страны).

$$\text{ОПСр} = \frac{\text{Показатель, характеризующий объект А}}{\text{Показатель, характеризующий объект Б}}$$

Таким образом, построение и совершенствование статистических показателей должно основываться на соблюдении двух основных принципов:

1) объективности и реальности (показатели должны правдиво и адекватно отражать сущность соответствующих экономических и социальных категорий (понятий));

2) всесторонней теоретической и методологической обоснованности (определение величины показателя, его измеримость и сопоставимость в динамике должны быть научно аргументированы, четко и доступно сформулированы и однозначно, в единообразном толковании применимы). Кроме того, величины показателей должны правильно количественно измеряться с учетом уровня, масштабов и качественных признаков состояния или развития соответствующего экономического или социального явления (отраслевой и региональный уровни, отдельное предприятие или работник и т. п.). При этом построение показателей должно носить сквозной характер, позволяющий не только суммировать соответствующие показатели, но и обеспечивать их качественную однородность в группах и совокупностях,

переход от одного показателя к другому для полной характеристики объема и структуры более сложной категории или явления. Наконец, построение статистического показателя, его структура и сущность должны предусматривать возможность всесторонне анализировать изучаемое явление или процесс, характеризовать особенности его развития, определять влияющие на него факторы.

Исчисление статистических величин и анализ данных об изучаемых явлениях – это третий и завершающий этап статистического исследования. В статистике рассматривают несколько видов статистических величин: абсолютные, относительные и средние величины. К числу обобщающих статистических показателей относятся также аналитические показатели рядов динамики, индексы и др.

Лекция 13. Средние величины

Цель: изучить виды среднеарифметических величин

Задачи: изучить методику расчета среднеарифметических величин.

План:

1. Виды среднеарифметических величин.
2. Механизм расчета.

Средние величины – это обобщающие показатели, выражающие типичные размеры количественно изменяющихся признаков качественно однородных явлений; характеризуют общий уровень этого признака, отнесенный к единице совокупности. Она применяется в случаях перехода от конкретных значений признака к характеристике всей совокупности.

В зависимости от характера изучаемых явлений, от конкретных задач и целей статистического исследования различают следующие виды средних:

- 1) среднеарифметическая;
- 2) среднегармоническая;
- 3) среднеантигармоническая;
- 4) среднестепенная;
- 5) среднеквадратическая;
- 6) распределительные средние:
 - мода;
 - медиана;
- 7) среднехронологическая.

1. Средняя арифметическая простая (\bar{X}) равна сумме произведений значений признака, деленной на их количество.

$$\bar{X} = \frac{\sum X_i}{n}$$

где $X_1, X_2 \dots X_n$ - значение признака;

n – число значений признака.

Например, известна сменная выработка рабочих бригады токарей:

Табельный номер рабочего	1	2	3	4	5
Количество изготовленных деталей, шт.	21	19	20	18	21

Требуется определить среднюю выработку бригады.

Для ее нахождения используется формула средней арифметической простой:

$$\bar{X} = \frac{\sum X_i}{n}$$

По приведенным данным необходимо определить среднюю выработку бригады. Для этого просуммируем количество всех изготовленных деталей (значение признака) и разделим на количество рабочих (число значений признака):

$$\bar{X} = \frac{21+19+20+18+21}{5} = 19,8 \approx 20 \text{ шт}$$

Среднеарифметическая взвешенная может быть определена двумя способами: *прямым* – отношением суммы произведений значений признака на их частоту, на сумму частот, и *способом «От нуля»*.

1. Прямым способом среднерифметическая взвешенная (\bar{x}) определяется по формуле:

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

где f_i – частота, т. е. число случаев возникновения того или иного значения признака.

Исходные данные для расчёта среднеарифметической взвешенной приведены в таблице 1

Таблица 1. Продажа акций на торгах условной фондовой биржи.

Сделка	Количество проданных акций шт. f_i	Курс продажи акций в руб. x_i
1	500	1080
2	300	1050
3	1100	1145

Определим *среднеарифметическую взвешенную* *прямым способом* по формуле:

$$\bar{X} = \frac{500 \cdot 1080 + 300 \cdot 1050 + 1100 \cdot 1145}{500 + 300 + 1100} = 1112,9 \text{ руб}$$

Следует помнить. ! При расчете средней арифметической по интервальному вариационному ряду для выполнения необходимых вычислений от интервалов переходят к их серединам.

Тогда зависимость для расчета среднеарифметической взвешенной имеет вид

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

где x_i - середина i –го интервала.

2. Способом «от нуля» среднеарифметическая может быть исчислена по формуле:

$$\bar{X} = X_0 \pm Z,$$

где X_0 – значение признака, принятого за «0». За «0» обычно принимается значение признака, расположенного в середине вариационного ряда (В нашем примере $X_0 = 25$ лет). Поэтому ряд необходимо стараться сделать нечётным. При чётном количестве значений ряда, за «0», может быть принято значение признака с наибольшей частотой.

\bar{Z} – условная среднеарифметическая, определяемая отношением суммы отклонений значений признаков от значения признака, принятого за «0», умноженной на соответствующие частоты, к сумме частот, по формуле:

$$\bar{Z} = \frac{\sum (x_i - x_0) f_i}{\sum f} = -3/20 = -0,15 \text{ руб.}$$

Приведём пример расчета среднеарифметической способом «от 0» (табл. 2).

Таблица 2. Пример исчисления среднеарифметической способом «от 0»

Возраст посетителей лет	Число посетителей	$X_i f_i$	Отклонение значений признака от условной средней (от 0)	$(X - X_0) \times f$
X	f		$X - X_0 = Z_i$	
5	2	10	-20	-4
15	5	75	-10	-5
25	8	200	0	0
35	4	140	10	+4
45	1	45	20	+2
Итого	20	470	-	-3

1. Подставляя в вышеприведенную формулу полученные значения, определим среднеарифметическую прямым способом:

$$\bar{X} = -0,15 \cdot 10 + 25 = 23,5 \text{ года,}$$

2. Рассчитаем среднеарифметическую Способом «от нуля».

За «0» принимаем значение признака, расположенного в середине вариационного ряда (В нашем примере $X_0 = 25$ лет).

$\bar{X} = 470/20 = 23,5$ года., то есть получили ту же величину, что и «прямым способом».

2. Средняя гармоническая имеет более сложную конструкцию, чем средняя арифметическая. Используется в тех случаях, когда статистическая информация не содержит частот по отдельным значениям признака, а представлена произведением значения признака на частоту.

Средняя гармоническая взвешенная используется в тех случаях, когда известен числитель исходного соотношения средней, но неизвестен знаменатель.

1), применяется среднегармоническая простая, определяемая по формуле:

$$\bar{X}_n = \frac{n}{\sum \frac{1}{x_i}}$$

Рассмотрим пример использования средней гармонической простой:

Три предприятия производят микроволновые печи. Себестоимость их производства на 1-ом предприятии составила 4000 руб., на 2-ом -3000 руб., на 3-ем – 5000 руб. Необходимо определить среднюю себестоимость производства микроволновой печи при условии, что на каждом предприятии общие затраты на ее изготовление составляют 600 тыс. руб.

Применять среднюю арифметическую в данном случае нельзя, так как предприятия выпускают разное количество

микроволновых печей:

первое – 150 шт. (600000/4000);

второе – 200 шт. (600000/3000);

третье – 120 шт. (600000/5000).

Среднюю себестоимость микроволновой печи можно получить, если общие затраты трех предприятий разделить на общий выпуск:

$$\overline{X}_n = \frac{600 + 600 + 600}{150 + 200 + 120} = 3,830 \text{ руб.}$$

2) Для исчисления среднегармонической Xf_i обозначим M_i , т.е.

$$Xf_i = M_i, \text{ отсюда } f_i = M_i/X_i.$$

Преобразуем формулу среднеарифметической – вместо Xf_i подставим M_i и получим среднеарифметическую взвешенную, исчисленную гармоническим способом:

$$\overline{X}_n = \frac{\sum M_i}{\sum \frac{1}{x_i} \times M}$$

В случаях, когда произведение Xf_i одинаково или равно единице ($M_i = 1$), применяется среднегармоническая простая, определяемая по формуле:

$$\overline{X}_n = \frac{n}{\sum \frac{1}{x_i}}$$

При работе со сгруппированными данными используется средняя гармоническая взвешенная:

Таблица 2 - Валовой сбор и урожайность подсолнечника по Центрально – Черноземному району (в хозяйствах всех категорий)

Область Валовой сбор, тыс. т	Валовой сбор, тыс. т	Урожайность, ц\га
Белгородская	97,0	16,1
Воронежская	204,0	9,5
Курская	0,5	4,8
Тамбовская	16,0	10,9
Липецкая	69,0	7,0

В общем случае средняя урожайность любой сельскохозяйственной культуры по нескольким территориям, агрофирмам, крестьянским хозяйствам и. т. п. может быть определена только на основе следующего исходного соотношения.

$$ИСС = \frac{\text{Общий валовой сбор, тыс.ц}}{\text{Общая посевная площадь, тыс.га.}}$$

Общий валовой сбор определяется суммированием валового сбора по областям. Однако данные о посевных площадях в явном виде в таблице отсутствуют. Их косвенно можно рассчитать разделив валовой сбор по каждой области на урожайность. Тогда, определим искомую среднюю, предварительно переведя тонны в центнеры

$$\bar{X}_n = \frac{970 + 2040 + 5 + 160 + 690}{\frac{970}{16,1} + \frac{2040}{9,5} + \frac{5}{4,8} + \frac{160}{10,9} + \frac{690}{7,0}} = 9,9 \text{ ц/ га}$$

3. Для характеристики некоторых процессов подходит еще одна своеобразного вида средняя, так называемая **антигармоническая**. Формула ее такова:

$$\bar{x} = \frac{\sum x_i^2 f_i}{\sum x_i f_i} \text{ (взвешенная)}$$

$$\bar{x} = \frac{\sum x_i^2}{\sum x_i} \text{ (невзвешенная)}$$

Приведем примеры ее использования.

1, Имеется n учителей, и каждый подготовил k_i , учеников. Если каждый ученик подготовит столько же учеников, сколько подготовил его учитель, то среднее соотношение учеников и учителей (т. е. как бы производительность работы учителя или темп распространения знаний) выразится средней антигармонической. Пусть 6 учителей подготовили соответственно 8, 7, 14, 23, 15, 11 учеников. Тогда

$$\bar{x} = \frac{8^2 + 7^2 + 14^2 + 23^2 + 15^2 + 11^2}{8 + 7 + 14 + 23 + 15 + 11} = 15,4$$

2. Имеется n отраслей, k_i , – эффективность вложений в отрасль i , т. е. 1 руб. вложений в текущем году дает доход k_i рублей в следующем году. Если k_i – неизменны, то эффективность вложений E выразится средней антигармонической, т. е.

$$E = \frac{\sum_{i=1}^n k_i^2}{\sum_{i=1}^n k_i} .$$

Допустим, имеются три отрасли. Пусть вложения в них имеют эффективность соответственно $k_1 = 1,1$; $k_2 = 1,2$ и $k_3 = 1,4$. Тогда

$$E = \frac{(1,1)^2 + (1,2)^2 + (1,4)^2}{1,1 + 1,2 + 1,4} = 1,246$$

3. Имеется некоторая совокупность людей. Среди них n женщин. Пусть k_i – число детей у женщины. Предположим, что каждая дочь имеет столько же детей, сколько имела ее мать. Предположим, что число девочек равно числу мальчиков. Тогда соотношение численности между поколениями выразится средней антигармонической:

$$T = 2 \times \frac{\sum_{i=1}^n \left(\frac{1}{2} k_i \right)^2}{\sum_{i=1}^n \frac{1}{2} k_i} = \frac{\sum_{i=1}^n k_i^2}{\sum_{i=1}^n k_i}$$

Пусть у десяти женщин соответственно по 1, 2, 3, 2, 3, 5, 1, 1, 2, 4 детей. Тогда

$$T = \frac{1^2 + 2^2 + 3^2 + 2^2 + 3^2 + 5^2 + 1^2 + 1^2 + 2^2 + 4^2}{1 + 2 + 3 + 2 + 3 + 5 + 1 + 1 + 2 + 4} = 3,083$$

Таким образом, на каждого человека первого поколения будет приходиться в среднем 3,083 человека следующего поколения.

Таким образом, выбор вида среднего показателя оказывает существенное влияние на его численную величину. Выбор вида средней определяется в каждом отдельном случае путем анализа исследуемой совокупности, изучения содержания явления.

4. Степенная средняя (x_c) используется в случаях, когда определяется среднее значение, выраженное функцией k -го порядка. Частным случаем ее является среднеквадратическая.

5. Среднеквадратическая $\overline{X^k}$ применяется в случае, когда осредняемое значение признака выражено в виде квадратных функций.

Степенная средняя и среднеквадратическая может быть простой и взвешенной и исчисляется соответственно по формулам:

$$\begin{aligned}\bar{x}_c &= \sqrt[k]{\frac{\sum x_i^k}{n}} & \bar{X}_k &= \sqrt[k]{\frac{\sum x_i^k}{n}} \\ \bar{x}_c &= \sqrt[k]{\frac{\sum x_i^k f_i}{\sum f_i}} & \bar{X}_k &= \sqrt[k]{\frac{\sum x_i^k f_i}{\sum f_i}}\end{aligned}$$

Имеется 10 квадратов с различной длиной сторон. Необходимо определить среднюю сторону одного квадрата (табл. 3).

Таблица 3. Исходные данные для расчета среднеквадратической. :

Сторона квадрата, см., (X)	Количество квадратов, (f)	X^2	$X^2 \times f_i$
4	1	16	16
6	3	36	108
8	5	64	320
10	1	100	100
Итого	10	-	544

Средняя сторона одного квадрата определится как среднеквадратическая взвешенная по формуле:

$$\bar{X}_K = \sqrt{\frac{544}{10}} = \sqrt{54,4} = 7,4 \text{ см.}$$

6. Распределительные средние включают моду и медиану.

Мода характеризует центр распределения по «весу» статистических единиц.

В дискретном ряду модой будет значение признака, частота которого наибольшая.

В интервальном ряду мода (M_0) находится в пределах того интервала, частота которого наибольшая.

Моду находим по формуле:

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

где :

x_o – начало модального интервала;

h_m – величина модального интервала;

f_m – частота модального интервала;

f_{m-1} – частота интервала, предшествующая модельному;

f_{m+1} – частота интервала, следующего за модальным.

Медиана – вариант, занимающий среднее место в вариационном ряду и делящий его на две равные части.

В дискретном ряду медианным вариантом будет вариант, расположенный в середине вариационного ряда и делящий его на две равные части. Для нахождения медианы необходимо предварительно определить номер медианного варианта (NM_e). В дискретном ряду с нечетной суммой частот, (NM_e) исчисляется следующим образом:

$$NM_e = \frac{\sum f_i}{2} + 0,5$$

где $\sum f_i$ – сумма частот.

При четном числе частот имеют место два медианных варианта (NM_e^1 , NM_e^2), номера которых определяются следующим образом:

$$NM_e^1 = \frac{\sum f_i}{2} \qquad NM_e^2 = \frac{\sum f_i}{2} + 1 = NM_e^1 + 1$$

Медианным вариантом будет тот вариант, с прибавлением частот которого сумма частот будет больше номера медианы. Следовательно, для определения медианного варианта необходимо последовательно суммировать частоты.

7. Средняя хронологическая - это средний уровень ряда динамики, т. е. средняя, исчисленная по совокупности значений показателя в разные моменты или периоды времени.

В зависимости от вида ряда динамики применяются различные способы ее расчета, а именно расчет средней хронологической интервального ряда и средней хронологической моментного ряда.

Средней хронологической интервального (более распространенного) ряда является средняя величина из уровней интервального ряда динамики, которая исчисляется по формуле:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

где \bar{y} - средний уровень ряда;

y_i - уровень ряда динамики;

n - число членов ряда

Лекция 14 .Группировка статистических данных

Цель: классифицировать статистические показатели.

Задачи: образовывать однородные группы на основе разделения статистической совокупности.

План:

1. Статистические группировки.
2. Принципы построения группировок.

Виды статистических группировок

Группировкой называется разделение множества единиц изучаемой совокупности на группы по отдельным существенным для них признакам.

В самом общем виде группировки можно классифицировать следующим

образом:



Статистическая группировка – это процесс образования однородных групп на основе расчленения статистической совокупности на части или объединения изучаемых единиц в частные совокупности по существенным для них признакам. Каждая из этих групп характеризуется системой статистических показателей. Например, группировка промышленных предприятия по формам собственности и т. д.

Особым видом группировок является классификация. Классификация – это как бы стандарт, в котором каждая единица совокупности может быть отнесена лишь к одной группе или подгруппе. Классификация основывается на самых существенных признаках, которые меняются очень мало

(например, классификация отраслей деятельности, классификация основных фондов и т. д.). Таким образом, классификация – это узаконенная, общепринятая, нормативная группировка.

Метод группировок применяется для решения трех основных задач. Во-первых, выделение социально-экономических типов явлений. Во-вторых, изучение структуры явления и структурных сдвигов, происходящих в нем. В-третьих, выявление взаимосвязей и взаимозависимостей между явлениями и признаками, характеризующими эти явления. В соответствии с этими задачами различают следующие виды статистических группировок: типологические, структурные и аналитические (факторные).

Типологическая группировка решает задачу выявления и характеристики социально-экономических типов (частных подсовокупностей) путем разделения качественно разнородной совокупности на классы, социально-экономические типы, однородные группы единиц в соответствии с правилами научной группировки.

Признаки, по которым производится распределение единиц изучаемой совокупности на группы, называются **группировочными признаками**, или основанием группировки.

При построении типологической группировки в качестве группировочных признаков могут выступать как атрибутивные, так и количественные признаки. Примером типологической группировки по атрибутивному признаку – это группировка предприятий по формам собственности, а по количественному признаку – группировка студентов по успеваемости (выделяются группы успевающих и неуспевающих студентов).

Структурной группировкой называется группировка, в которой происходит разделение выделенных с помощью типологической группировки типов явлений, однородных совокупностей на группы, характеризующие их структуру по какому-либо варьирующему признаку. В качестве группировочных признаков могут рассматриваться атрибутивные и количественные признаки.

При группировке по атрибутивному признаку группы отличаются друг от друга по характеру признака. Число групп, на которые делится изучаемая совокупность, определяется числом градаций атрибутивного признака.

Одной из основных задач статистических группировок состоит в исследовании связей и зависимостей между признаками единиц статистической совокупности, которая решается с помощью построения аналитических группировок. **Аналитическая группировка** – это группировка, выявляющая взаимосвязи взаимозависимости между изучаемыми социально-экономическими явлениями и признаками их характеризующими. Особенности аналитической группировки состоит в том, что единицы совокупности группируются по факторному признаку, а расчет групповых средних производится по значениям результативного признака. То есть, каждая выделенная группа характеризуется средними величинами результативного признака. По изменению этих величин и определяется наличие связей и зависимостей между признаками.

Факторными называются признаки, оказывающие влияние на изменение результативных признаков. **Результативными** называются признаки, изменяющиеся под влиянием факторных признаков.

В зависимости от степени сложности массового явления и от задач анализа группировки могут производиться по одному или нескольким признакам.

Если группы образуются по одному признаку, группировка называется простой (например, распределение населения по возрастным группам, а семей – по уровню доходов и т. д.).

Группировка по двум или нескольким признакам называется сложной.

Если группы, образованные по одному признаку, делятся на подгруппы по второму, а последние – на подгруппы по третьему и т. д. признакам, т. е. в основании группировки лежит несколько признаков, взятых в комбинации, то такая группировка называется **комбинационной** (например, дополнив простую группировку населения по возрастным группам группировкой по

полу, получим комбинационную группировку). Комбинационные группировки позволяют изучить единицы совокупности одновременно по нескольким признакам.

Если группировка строится по атрибутивному признаку, то групп, как правило, будет столько, сколько имеется градаций, видов состояний у этого признака. Например, группировка предприятий по формам собственности учитывает муниципальную, федеральную, собственность субъекта Федерации и частную.

Если группировка проводится по количественному признаку, то тогда необходимо обратить особое внимание на число исследуемых объектов и степень колеблемости группировочного признака. Если количественный группировочный признак меняется **прерывно** (дискретно), т. е. может принимать некоторые – чаще целые значения (например, тарифный разряд рабочих), то число групп должно соответствовать количеству значений признака.

Принципы построения группировок

Признаки, по которым проводится группировка, называют группировочными признаками. Группировочные признаки могут иметь количественное (объем, доход, и т.д.) и качественное выражение (пол человека, отраслевая принадлежность, семейное положение и т.д.). Число групп определяется числом интервалов, на которые разбивается весь диапазон изменения признака.



Каждый интервал имеет нижнюю и верхнюю границы или одну из них. Если у интервала указана лишь одна граница, то это открытый интервал, а

если обе – то закрытый. Закрытые интервалы подразделяются на равные и неравные, а также специализированные и произвольные. Если можно заранее установить определенное количество групп, то величину равного интервала можно вычислить по формуле

$$i = \frac{X_{\max} - X_{\min}}{n}$$

где i - величина равного интервала; X_{\max} - X_{\min} - амплитуда колебания признака; n - число групп.

При определении количество групп необходимо стремиться к тому, чтобы были учтены особенности изучаемого явления. Поэтому число групп должно быть оптимальным, в каждой группе должно входить достаточно большое число единиц совокупности, что отвечает требованию закона больших чисел.

При большом количестве наблюдений количество групп k – определяют по формуле Стерджесса:

$$k = 1 + 3,322 \times \lg n$$

где n – число единиц совокупности, в общем, ее объеме.

Результат при таком расчете округляют до целого числа.

Интервалы могут быть равные и неравные. Последние делятся прогрессивно – возрастающие и прогрессивно – убывающие.

Группировка с равными интервалами целесообразно в тех случаях, когда вариация проявляется в сравнительно узких границах и распределение является практически равномерным (например, при группировке рабочих одной профессий по размеру заработной платы, посевов какой – либо культуры по урожайности).

Для группировок с равными интервалами величина интервала определяется по формуле:

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

где x_{\max} , x_{\min} – наибольшее и наименьшее значения признака.

Если в результате деления в формуле получится нецелое число и возникает необходимость в округлении, то округлять нужно, в большую сторону, а не в меньшую.

Пусть имеем группировку рабочих по уровню заработной платы (руб.):
6000 – 6300; 6300 – 6600; 6600 – 6900; 6900 – 7200; 7200 – 7500.

В этом распределении имеет место неопределенность: к какой группе, например, отнести рабочего с заработком в 6300 руб., к первой или второй? Для устранения неопределенности используется правило. Если величина признака единицы совпадает с верхней границей группы, то эта единица переходит к следующей группе, исключая последнюю группу. Значит рабочий, получающий 6300 руб., должен быть отнесен ко второй группе. Если же заработная плата рабочего равна 7500 руб., то этого рабочего следует отнести к последней группе.

Если размах вариации ($R = x_{\max} - x_{\min}$) признака статистической совокупности велик и значения признака варьируются неравномерно, то необходимо использовать группировку с **неравными интервалами**.

Статистическая таблица

Результаты группировочного материала оформляются в виде таблиц, где он излагается в наглядно-рациональной форме. Не всякая таблица может быть статистической. Табличные формы календарей, тестовых и опросных листов, таблица умножения не являются статистическими.

Статистическая таблица - это цифровое выражение итоговой характеристики всей наблюдаемой совокупности или ее составных частей по одному или нескольким существенным признакам. Статистическая таблица содержит два элемента: подлежащее и сказуемое. Подлежащее – объект, характеризующийся цифрами. Обычно подлежащее дается в левой части, в наименовании строк. Сказуемое - это цифровые показатели, с помощью которых дается характеристика выделенных в подлежащем групп и единиц.

Сказуемое формирует верхние заголовки. Различают простые, групповые и комбинационные таблицы. Часто результаты группировки представляются в графическом виде – с помощью гистограмм.

Лекция 15 Выборочное наблюдение. Причины применения выборочного наблюдения. Основная задача выборочного наблюдения.

Цель: грамотно интерпретировать полученные результаты, изучить сущность и значение.

Задачи: правильно решать практические задачи по данной теме

План:

1. Виды статистического наблюдения
2. Основные задачи выборочного наблюдения

Под **выборочным** понимается метод статистического исследования, при котором обобщающие показатели изучаемой совокупности устанавливаются по некоторой ее части на основе положений случайного отбора. Подлежащую изучению совокупность, из которой производится отбор части единиц, называют **генеральной совокупностью**. Отобранная из генеральной совокупности некоторая часть единиц, подвергающаяся обследованию, называется **выборочной совокупностью** или **выборкой**. Генеральная совокупность может быть конечной (число наблюдений $N = \text{const}$) или бесконечной ($N = \infty$), а выборка из генеральной совокупности – это всегда результат ограниченного ряда n наблюдений.

Долей выборки kn называется отношение числа единиц выборочной совокупности к числу единиц генеральной совокупности:

$$kn = n/N.$$

Значение выборочного метода состоит в том, что при минимальной численности обследуемых единиц проведение исследования осуществляется в более короткие сроки и с минимальными затратами труда и средств. Это повышает оперативность статистической информации, уменьшает ошибки регистрации.

В проведении ряда исследования выборочный метод является единственно возможным, например, при контроле качества продукции (товара, услуги), если проверка сопровождается уничтожением или разложением на составные части обследуемых образцов.

При соблюдении правил научной организации обследования выборочный метод даёт достаточно точные результаты, поэтому его целесообразно применять для проверки данных сплошного учёта. Минимальная численность обследуемых единиц позволяет провести исследование более тщательно и квалифицированно. Так, при переписи населения практикуются выборочные контрольные обходы для проверки правильности записей сплошного наблюдения.

Таким образом, основными причинами применения выборочного исследования являются:

- ◆ экономия;
- ◆ невозможность проведения сплошного исследования.

По сравнению с другими статистическими методами, применяющими не сплошное наблюдение, выборочный метод имеет важную особенность, которая заключается в том, что в основе отбора единиц для обследования положены принципы равных возможностей попадания в выборку каждой единицы генеральной совокупности. Именно в результате соблюдения этих принципов исключается образование выборочной совокупности только за счёт лучших или худших образцов. Это предупреждает появление систематических (тенденциозных) ошибок и делает возможным производить количественную оценку ошибки представительства (репрезентативности).

Проведение исследования социально-экономических явлений выборочным методом складывается из ряда последовательных этапов:

- 1) обоснование (в соответствии с задачами исследования) целесообразности применения выборочного метода исследования;
- 2) составление программы проведения статистического исследования выборочным методом;

- 3) решение организационных вопросов сбора и обработки исходной информации;
- 4) установление доли выборки, т.е. части подлежащих обследованию единиц генеральной совокупности;
- 5) обоснование способов формирования выборочной совокупности;
- 6) осуществление отбора единиц из генеральной совокупности для их обследования;
- 7) фиксация в отобранных единицах (пробах) изучаемых признаков;
- 8) статистическая обработка полученной в выборке информации с определением обобщающих характеристик изучаемых признаков;
- 9) определение количественной оценки ошибки выборки;
- 10) распространение обобщающих выборочных характеристик на генеральную совокупность. *При выборочном наблюдении дело имеет с двумя категориями обобщающих показателей с относительными и средними величинами.*

Относительные величины применяют для сводной характеристики совокупностей по альтернативному признаку; такая характеристика даётся в виде доли (удельного веса) тех единиц совокупности, которые обладают интересующим исследователя признаком. Например, при анализе качества продукции определяют относительную долю тех единиц, которые не выдерживают установленного стандарта качества, т.е. относятся к браку и т.д.

Во всех случаях, когда речь идёт о вариации альтернативных признаков, мы будем иметь дело с обобщающим показателем в виде относительной доли единиц. В генеральной совокупности доля единиц, обладающих изучаемым признаком, называется **генеральной долей**.

Кроме измерения доли, перед выборочным наблюдением может стоять задача измерения *среднего значения варьирующего признака во всей совокупности*. В этом случае имеют дело с признаками, вариация которых проявляется в разных количественных значениях у отдельных единиц

совокупности. Средняя величина изучаемого варьирующего признака - генеральной средней.

В выборочной совокупности долю изучаемого признака называют выборочной долей, или частью, а среднюю величину в выборке - выборочной средней. Основная задача выборочного исследования в сфере обслуживания состоит в том, чтобы на основе характеристик выборочной совокупности получить достоверные суждения о показателях доли или средней в генеральной совокупности.

Условия проведения выборки

Во-первых, она должна быть достаточно многочисленной, чтобы в ней могли проявиться закономерности, существующие в генеральной совокупности.

Во-вторых, элементы выборки должны быть отобраны объективно, независимо от воли исследователя, чтобы каждый из них имел одинаковые шансы быть отобранным или чтобы эти шансы были известны исследователю.

Пример 1. Имеются данные о зарплате рабочих в тыс.руб.:

Группы по з/пл. тыс. руб.	ГС - человек	Из них попали в выборку
10-13	100	5
13-16	150	10
16-19	400	30
19-22	200	45
22-25	150	10
Итого	1000	100

- Как видим, зарплату от 100 до 130 в ГС получают 10%, в ВС – 5%. Доля этой группы в ВС ниже, чем в ГС, ВС неточно представляет ГС.
- Зарплату от 190 до 220 в ГС получают 20%, а в выборку получающих такую зарплату попало 45%. Снова налицо проблема репрезентативности.

Виды выборки:

1. По участию единиц наблюдения в отборе:

- Повторная выборка, при которой попавшая в выборку единица после регистрации возвращается в генеральную совокупность и таким образом сохраняет равную возможность наряду с другими единицами быть использованной в дальнейшей процедуре отбора; при этом численность единиц генеральной совокупности N остается неизменной.

- Бесповторная выборка, при которой попавшая в выборку единица не возвращается в исходную совокупность и в дальнейшем выборе не участвует; при этом численность единиц генеральной совокупности N сокращается в процессе отбора.

2. По количеству ступеней отбора:

- Одноступенчатая, при которой каждая отобранная единица сразу подвергается изучению по данному признаку;

- Многоступенчатая, при которой производят отбор из генеральной совокупности отдельных групп, из которых выбирают подгруппы и т. д., на последней стадии единица отбора совпадет с единицей наблюдения.

Лекция 16. Статистическое изучение динамики. Анализ динамики

Социально – экономические явления общественной жизни находятся в непрерывном развитии. Их изменение во времени статистика изучает при помощи построения и анализа рядов динамики.

Начиная изучение данной темы, необходимо обратить внимание на классификацию рядов динамики, различия между ними, так как отнесение ряда динамики к тому или иному виду имеет важное значение для их изучения. Выбор соответствующих приёмов и способов анализа определяется характером исходных данных и зависит от задач исследования.

Данная тема знакомит студентов с задачами, решение которых дает возможность усвоить правила построения и анализа рядов динамики для характеристики изменения социально – экономических явлений во времени, выявления основной тенденции, закономерностей их развития.

Ряд динамики – числовые значения статистического показателя, представленные во временной последовательности. Он состоит из двух граф: в первой указываются периоды (или даты), во второй – показатели, характеризующие изучаемый объект за эти периоды (или на эти даты).

Показатели второй графы носят название *уровни ряда*: первый показатель называется *начальным* уровнем, последний – *конечным*. Уровни рядов динамики могут быть выражены абсолютными, средними или относительными величинами. Исходя из этого, ряды динамики подразделяются на **ряды абсолютных, относительных и средних величин**.

Ряды динамики могут быть двух видов: интервальные и моментные.

В **интервальном ряду** приводятся данные, характеризующие величину показателя за определенные периоды (сутки, месяц, квартал, год и т. д.). Особенность интервальных рядов из абсолютных величин является то, что их уровни можно суммировать, получая новые значения *объема явления*, относящиеся к более длительным периодам.

В **моментном ряду** динамики приводятся данные, характеризующие

размеры явления на определенные моменты (даты) времени. Уровни моментных рядов динамики суммировать нельзя; сумма не имеет экономического смысла, так как каждый последующий уровень полностью или частично включает в себя предыдущий уровень. Однако разность уровней имеет смысл, характеризуя увеличение или уменьшение уровня ряда между датами учета.

Важнейшим условием правильного формирования рядов динамики является **сопоставимость уровней, образующих ряд**. Основным требованием сопоставимости уровней является одинаковая методология их исчисления для всех периодов или дат. При этом все уровни должны быть даны не только в одинаковых, но и равноценных единицах измерения. Условием сопоставимости данных является также одинаковая полнота охвата различных частей явления, представленного рядом динамики. Уровни показателей в интервальных динамических рядах должны относиться к периодам с одинаковой продолжительностью. Для моментных рядов должна соблюдаться неизменность даты учета.

Вопрос о том, следует ли считать условием сопоставимости данных динамического ряда одинаковость границ территории, к которой относятся данные, решается по – разному. Если ставится задача изучения изменения явления в связи с изменением территории, то в этом случае сопоставляются данные, относящиеся к различной территории. Если же ставится задача изменения темпов развития явления, то сравниваемые показатели должны относиться к неизменной территории. Поэтому, прежде чем анализировать ряд динамики, необходимо обеспечить сопоставимость уровней ряда. Для этого выполняется дополнительные расчеты, которые называются **смыканием уровней динамических рядов**.

Условием же сопоставимости уровней интервального временного ряда является равенство периодов, за которые приводятся данные. Если это условие нарушено, то исследуемый ряд подвергают дополнительной обработке - рассчитывают уровни ряда в среднем на единицу времени.

Пример 1. Объем инвестиций по фирме характеризуется следующими данными .

Таблица 1

	Периоды			
	2000-2005	2006-2009	2010-2016	2016
Объем капитальных вложений (в сопоставимых ценах, млн. руб.	1140,8	1225,5	2960,6	508,8

Уровни представленного ряда не сопоставимы между собой, так как показатели относятся к периодам с различной продолжительностью. Для того, чтобы выявить изменение динамики капитальных вложений во времени, необходимо определить величину капитальных вложений на одну и ту же единицу каждого периода – один год.

Объем капитальных вложений за один год составляет (млн. руб.):

2000 – 2005 гг. – 228,16 (1140,8 : 5);

2006 – 2009 гг. – 408,50 (1225,5 : 3);

2010 – 2016 гг. – 493,43 (2960,6 : 6);

2016 г. – 508,80.

Из анализа полученных данных видно, что объем капитальных вложений по фирме за период 2000 – 2016 гг. имеет тренд к повышению.

Если несопоставимость в рядах динамики вызвана административно - территориальными изменениями, то для изучения развития явления необходимо построить ряд сопоставимых уровней в новых территориальных границах.

Для анализа рядов динамики применяется ряд показателей, которые определяют изменение общественных явлений во времени. К ним относятся: уровень ряда, абсолютный прирост, темпы роста и прироста, абсолютное

значение одного процента прироста, средний абсолютный прирост, средние темпы роста и прироста.

Абсолютный прирост – показатель, который характеризует абсолютный размер увеличения или уменьшения текущего уровня ряда по сравнению с начальным (базисным) или предшествующим (цепным):

$$\Delta \delta = Y_m - Y_{\delta};$$

$$\Delta \zeta = Y_m - Y_{m-1}$$

где $\Delta \delta$ – абсолютный прирост базисный; $\Delta \zeta$ – абсолютный прирост цепной; Y_m – уровень текущий; Y_{δ} – уровень базисный; Y_{m-1} – уровень предшествующий.

Средний абсолютный прирост характеризует средний ежегодный прирост уровня изучаемого явления. Исчисляется по формуле

$$\overline{\Delta} = \frac{Y_n - Y_{\delta}}{n - 1},$$

где n – число уровней ряда; Y_n – последний уровень ряда.

Темп роста – показатель, который характеризует, во сколько раз данный уровень ряда больше начального (базисного) или предшествующего (цепного) уровня или какую долю от них составляет. Исчисляется он по формуле:

$$T_{\delta}^{\delta} = \frac{Y_T}{Y_{\delta}};$$

$$T_{\zeta}^{\zeta} = \frac{Y_T}{Y_{T-1}}$$

где T_{δ}^{δ} – темп роста базисный; T_{ζ}^{ζ} – темп роста цепной.

Средний темп роста $\overline{T_p}$ показывает, во сколько раз в среднем увеличивается текущий уровень изучаемого явления по сравнению с уровнем предшествующего периода за изучаемый период.

Исчисляется он по формуле:

$$\bar{T}_p = \sqrt[n-1]{\frac{Y_n}{Y_0}}.$$

Темп прироста – показатель, который характеризует относительную величину прироста по отношению к начальному (базисному) или предшествующему (цепному) уровню. Исчисляется он по формуле

$$T_{np}^{\bar{}} = \frac{Y_T - Y_0}{Y_0};$$

$$T_{np}^u = \frac{Y_T - Y_{T-1}}{Y_{T-1}}$$

где $T_{np}^{\bar{}}$ – темп прироста базисный; T_{np}^u – темп прироста цепной.

Между темпами роста и прироста существует взаимосвязь. Темп прироста равен темпу роста минус единица или 100%. Средний темп прироста \bar{T}_{np} показывает, на сколько процентов увеличился или уменьшился в среднем текущий уровень по сравнению с предшествующим.

$$T_{np}^u = \frac{Y_T}{Y_{T-1}} - 1$$

Абсолютное значение одного процента прироста $A(\%)$ показывает сколько абсолютных единиц скрывается за каждым процентом прироста. Исчисляется оно по формуле:

$$A\% = \frac{\Delta_i}{T_{np}},$$

где Δ_i – абсолютный прирост;

T_{np} – соответствующий темп прироста

Разложение рядов динамики

Уровни любого ряда динамики формируются под совместным влиянием факторов, различных как по характеру, так и силе воздействия. В первую очередь необходимо выделить факторы эволюционного характера, оказывающие постоянное воздействие и определяющие общее направление

развития явления, его долговременную эволюцию. Такие изменения динамического ряда называют основной тенденцией развития или трендом.

Вторую группу факторов составляют факторы осциллятивного характера, оказывающие периодическое воздействие.

Они вызывают *циклические и сезонные колебания* уровней динамического ряда.

Циклические (или периодические) долговременные колебания – *это регулярные колебания, вызываемые постоянно действующими причинами*, например, циклы экономической конъюнктуры, циклы гарвардской школы. Схематично циклические колебания можно представить в виде синусоиды $y = \sin t$ (значение признака вначале возрастает, достигает определенного max, затем снижается, достигает своего min, вновь возрастает и т.д.).

Колеблемость рядов динамики характеризуется размахом (амплитудой) колебания (R_t), средним линейным отклонением ($\overline{d_t}$) и дисперсией (σ_t^2), которые *исчисляются по тренду* (A), который равен разности между абсолютными значениями ряда и выравненными по уравнению прямой линии; т. е. $\Delta = y_i - \overline{y}_i$.

$$R_t = \Delta_{\max} - \Delta_{\min}.$$

$$\overline{d_t} = \frac{1}{n} \sum \Delta_i,$$

$$\sigma_t^2 = \frac{1}{n} \sum \Delta_i^2.$$

Относительным показателем колеблемости уровня динамического ряда является коэффициент колеблемости, который исчисляется по формуле

$$V = \frac{\sigma_t}{\overline{y}}.$$

Мерой устойчивости ряда динамики (K_{yc}) является величина, исчисленная согласно формуле:

$$K_{yc} = 1 - V.$$

Выявление тренда

Первая задача, которая возникает при анализе рядов динамики, заключается в выявлении и описании основной тенденции развития изучаемого явления (тренда).

Трендом называется плавное и устойчивое изменение уровней явления во времени, свободное от случайных колебаний.

Изучение тренда включает в себя два этапа:

1. Проверка ряда на наличие тренда;
2. Выравнивание ряда динамики и непосредственное выделение тренда

Проверка ряда на наличие тренда проводится разными методами, самым простым из которых является метод средних. Суть его заключается в следующем: изучаемый ряд динамики разбивается на несколько интервалов (чаще всего на два), для каждого из которых определяется средняя величина \bar{y}_1 и \bar{y}_2 . Выдвигается гипотеза о существенном различии средних. Если выдвинутая гипотеза принимается, то признается наличие тренда.

Для непосредственного выявления тренда используют следующие методы:

- метод укрупнения интервалов;
- метод скользящей средней;
- метод аналитического выравнивания.

Все перечисленные методы относятся к группе методов сглаживания, предполагающих наличие в исходном ряду динамики только одной компоненты – тренда. Метод укрупнения интервалов является одним из наиболее простых методов непосредственного выявления основной тенденции. При использовании этого метода *ряд динамики, состоящий из мелких интервалов, заменяется рядом, состоящим из более крупных интервалов.*

Так как на каждый уровень исходного ряда влияют факторы, вызывающие их разнонаправленное изменение, то это мешает видеть

основную тенденцию. При укрупнении интервалов влияние факторов нивелируется, и основная тенденция проявляется более отчетливо. *Расчет среднего значения уровня по укрупненному интервалу осуществляется по формуле простой средней арифметической.*

Недостатком этого способа является то, что сокращается число уровней ряда, а это не позволяет учитывать изменения внутри укрупненного интервала. К его *преимуществам* можно отнести сохранение природы явления.

ЛЕКЦИЯ 17 ИНДЕКСЫ И ИХ ПРИМЕНЕНИЕ В ЭКОНОМИЧЕСКОМ АНАЛИЗЕ

Цель: характеристика общего изменения сложного социально - экономического показателя и отдельных его элементов.

Задачи: измерение влияния факторов на общую динамику сложного показателя, включая характеристику влияния изменения структуры явления.

План:

1. Индексные показатели

2. Методы расчета

Индекс – относительная величина, характеризующая изменение уровней сложных социально – экономических показателей во времени, в пространстве или по сравнению с некоторым эталоном (плановым, нормативным уровнем и т. п.). Если в качестве базы сравнения используется уровень за какой – либо предшествующий период, то получают индекс динамики; если же базой сравнения является уровень того же явления по другой территории, то получают территориальные индексы.

Индексные показатели вычисляются на высшей ступени статистического обобщения. Поэтому изучение данной темы должно базироваться на знании таких предшествующих тем как: «Сводка и группировка данных»; «Статистические показатели» и «Статистическое изучение динамики». С помощью индексного метода решаются следующие задачи:

1. характеристика общего изменения сложного социально - экономического показателя и отдельных его элементов;

2. измерение влияния факторов на общую динамику сложного показателя, включая характеристику влияния изменения структуры явления.

Индекс является результатом сравнения двух одноименных показателей, поэтому при их вычислении различают сравниваемый уровень (числитель индексного отношения), называемый **текущим** или **отчетным**, и

уровень, с которым производится сравнение знаменатель индексного отношения), называемый **базисным**.

Простейшим показателем, используемым в индексном анализе, является индивидуальный индекс, который характеризует изменение во времени (или в пространстве) отдельных элементов той или иной совокупности. Так, **индивидуальный индекс цены** рассчитывается по формуле:

$$i_p = \frac{P_1}{P_0},$$

где P_1 – цена товара в текущем периоде;

P_0 – цена товара в базисном периоде.

Например, если цена товара А в текущем периоде составляла 90 руб., а в базисном 75 руб., то индивидуальный индекс цены: $i_p = \frac{90}{75} = 1.2$, или 120,0%.

В данном примере цена товара А возросла по сравнению с базисным уровнем в 1,2 раза, или на 20 %.

Оценить изменение объемов продажи товара в натуральных единицах измерения позволяет **индивидуальный индекс физического объема реализации**:

$$i_q = \frac{q_1}{q_0},$$

где q_1 – количество товара, реализованное в текущем периоде;

q_0 – количество товара, реализованное в базисном периоде.

Изменение объема реализации товара в стоимостном выражении отражает **индивидуальный индекс товарооборота**:

$$i_Q = \frac{P_1 q_1}{P_0 q_0}.$$

Индивидуальные индексы, в сущности, представляют собой относительные показатели динамики или темпы роста и по данным за несколько периодов времени могут рассчитываться в цепной или базисной формах.

Агрегатный индекс - это сложный относительный показатель, который характеризует среднее изменение социально-экономического явления, состоящего из непосредственно несоизмеримых элементов. Исходной формой сводного индекса является агрегатная.

При расчете агрегатного индекса для разнородной совокупности находят такой общий показатель, в котором можно объединить все ее элементы. Рассмотрим пример с розничными ценами. Цены различных товаров, реализуемых в розничной торговле, складывать неправомерно, однако с экономической точки зрения вполне допустимо суммировать товарооборот по этим товарам. Если мы сравним товарооборот в текущем периоде с его величиной в базисном периоде, то получим **агрегатный индекс товарооборота**:

$$I_Q = \frac{\sum p_1 q_1}{\sum p_0 q_0}.$$

На величину данного индекса оказывают влияние как изменение цен на товары, так и изменение объемов их реализации. Для того чтобы оценить изменение только цен (индексируемой величины), необходимо количество проданных товаров (веса индекса) зафиксировать на каком-либо постоянном уровне. При исследовании динамики таких показателей, как цена, себестоимость, производительность труда, урожайность, количественный показатель обычно фиксируют на уровне текущего периода. Таким способом получают **агрегатный индекс цен** (по методу Пааше):

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}.$$

Числитель данного индекса содержит фактический товарооборот текущего периода. Знаменатель же представляет собой условную величину, показывающую, каким был бы товарооборот в текущем периоде при условии сохранения цен на базисном уровне. Поэтому соотношение этих двух категорий и отражает имевшее место изменение цен.

Числитель и знаменатель сводного индекса цен можно интерпретировать с точки зрения потребителей. Числитель представляет собой сумму денег, фактически уплаченных покупателями за приобретённые в текущем периоде товары. Знаменатель же показывает, какую сумму покупатели заплатили бы за те же товары, если бы цены не изменились.

Разность числителя и знаменателя будет отражать величину экономии (если знак «-») или перерасхода («+») покупателей от изменения цен.

Третьим индексом в данной индексной системе является **агрегатный индекс физического объема реализации** (по методу Ласпейреса). Он характеризует изменение количества проданных товаров не в денежных, а в физических единицах измерения:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}.$$

Весами в данном индексе выступают цены, которые фиксируются на базисном уровне.

Между рассчитанными индексами существует следующая взаимосвязь:

$$I_p \cdot I_q = I_{pq}.$$

Общий индекс цен может быть построен и как средняя геометрическая из агрегатных индексов цен Ласпейреса и Пааше, т.е. по формуле Фишера:

$$I_{цен} = \sqrt{\frac{\sum q_0 p_1}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_1 p_0}}.$$

Это так называемый индекс Фишера, рекомендуемый его автором в тех случаях, когда трудно отдать предпочтение весам q_0 или q_1 . Поскольку в этой формуле учтены веса обоих периодов, Фишер считал этот индекс идеальным.

Следует также обратить внимание на то, что если строится ряд индексов, то они могут быть построены или как цепные (ряд индексов, каждый из которых построен по отношению к предыдущему периоду), или как базисные (ряд индексов, построенных в сравнении с одной и той же базой). Произведение цепных индексов дает базисный индекс. Путем деления двух базисных индексов легко получить цепной индекс.

Задача 1. Имеются следующие данные о реализации плодово-ягодной продукции в области (табл. 1).

Таблица 1 - Реализация плодово-ягодной продукции в городе

Наименование товара	Июль		Август		Расчетные графы, руб.		
	цена за 1 кг, руб. p_0	продано, т q_0	Цена за 1 кг, руб. p_1	Продано, т q_1	p_0q_0	p_1q_1	p_0q_1
Черешня	150	24	140	21	3600	2940	3150
Персики	120	28	110	33	3360	3630	3960
Виноград	100	26	90	25	2600	2250	2500
Итого	-	-	-	-	9560	8820	9610

Определите:

- 1) общий индекс товарооборота;
- 2) общий индекс физического объема товарооборота;
- 3) общий индекс цен;
- 4) прирост товарооборота - всего, в том числе за счет изменения цен и объема продажи товаров.

Покажите связь между исчисленными индексами.

Решение:

1. Общий индекс товарооборота исчисляется по формуле:

$$I_Q = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{8820}{9560} = 0,922, \text{ или } 92,2\%.$$

Товарооборот в августе снизился на 7,8%, по сравнению с июлем.

2. Общий индекс физического объема товарооборота (количества проданных товаров) исчисляется по формуле:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{9610}{9560} = 1,005.$$

Это значит, что количество проданного товара в августе было больше на 0,5%, чем в июле.

3. Общий индекс цен равен:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{8820}{9610} = 0,918, \text{ или } 91,8\%.$$

т.е. цены на все товары в среднем снизились на 8.8%.

4. Прирост или снижение товарооборота исчисляется как разница между числителем и знаменателем индекса товарооборота:

$$E_Q = \sum p_1 q_1 - \sum p_0 q_0 = 8820 - 9560 = -740 \text{ тыс. руб.}$$

Это снижение обусловлено изменением цен на товары и изменением количества проданных товаров.

Снижение за счет изменения цен составил:

$$E_p = \sum p_1 q_1 - \sum p_0 q_1 = 8820 - 9610 = -790 \text{ тыс. руб.},$$

снижение за счет изменения количества проданных товаров:

$$E_q = \sum q_1 p_0 - \sum q_0 p_0 = 9610 - 9560 = 50 \text{ тыс. руб.}$$

Следовательно, снижение товарооборота на 740 тыс. руб. произошло за счет повышения количества проданных товаров на 50 тыс. руб. и за счет снижения цен на 790 тыс. руб. $[(-790) + (+50) = -740 \text{ тыс. руб.}]$.

Между исчисленными индексами существует связь:

$$I_Q = I_q \cdot I_p = 1,005 \cdot 0,918 = 0,922.$$

Числитель и знаменатель агрегатного индекса цен можно интерпретировать с точки зрения потребителей. Числитель представляет собой сумму денег, фактически уплаченных покупателями за приобретённые в текущем периоде товары. Знаменатель же показывает, какую сумму покупатели заплатили бы за те же товары, если бы цены не изменились.

Разность числителя и знаменателя будет отражать величину экономии (если знак «-») или перерасхода («+») покупателей от изменения цен:

$$E_{pq} = \sum p_1 q_1 - \sum p_0 q_1 = 8820 - 9610 = -790 \text{ тыс. руб.}$$

Индекс физического объема реализации составит:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{9610}{9560} = 1,005, \text{ или } 100,5\%.$$

Физический объем реализации (товарооборота) увеличился на 8,6%.

Используя взаимосвязь индексов, проверим правильность вычислений:

$$I_{pq} = I_p \cdot I_q = 0,9181 \cdot 1,005 = 0,922, \text{ или } 92,2\%.$$

Мы рассмотрели применение агрегатных индексов в анализе товарооборота и цен. При анализе результатов производственной деятельности промышленного предприятия приведенные выше сводные индексы соответственно называются индексом стоимости продукции, индексом оптовых цен и индексом физического объема продукции.

Лекция 18 Понятие структуры явления и ее виды

Цель: рассмотреть показатели для учета различия (подобия) структур.

Задачи: изучить сложные структуры, образуемые при последовательном дроблении системы на однородные группы элементов.

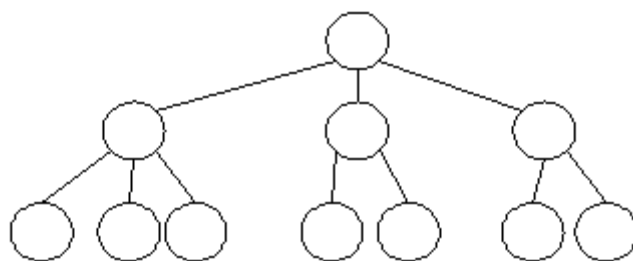
План:

1. Виды структуры.
2. Методы анализа структуры.

Развитие статистической совокупности проявляется не только в количественном росте или уменьшении элементов этой системы, но также и в изменении ее структуры. Понятие **структуры** очень тесно переплетается с понятием **группировка и классификация**. **Структура** — это строение, форма организации системы, состоящей из отдельных элементов и связей между ними.



Иерархической (древовидной) структурой называется сложная структура, образуемая при последовательном дроблении системы на все более однородные группы элементов. Она состоит из нескольких уровней («шагов» дробления).



Основное **преимущество** иерархической структуры заключается в ее большой информационной емкости (класс, подкласс, группа, подгруппа, вид, (разновидности), традиционности и привычности применения, хорошей приспособленности для различной обработки информации, а также возможности для создания при кодировании объектов классификации кодов, несущих смысловую нагрузку.

Недостатком является слабая гибкость ее структуры и заранее установленный порядок ступеней распределения, не допускающий включения при отсутствии резервной емкости новых объектов классификационных группировок и признаков. Вследствие этого изменение хотя бы одного признака ведет к перераспределению многих классификационных группировок. Иерархическая структура характеризуется не только долями объема признака, но и дополнительными показателями.

1. Характеристикой степени сложности структуры, а именно числом уровней дробления («порядок» структуры).

2. Средним порядком структуры, т.е. средним номером уровня, взвешенным по долям объема признака, дробление которых завершилось на данном уровне. Эта величина характеризует среднее число дроблений объема признака.

3. Общим числом конечных (т.е. не дробящихся далее) ветвей структуры.

4. Средним числом конечных ветвей, приходящихся на один уровень.

Баланс (фр. *balance* — буквально весы, равновесие) — это особая форма сопоставления структуры одной и той же величины признака, характеризуемой с двух разных сторон или в двух различных аспектах. В наиболее общей форме динамический баланс состоит из четырех составляющих: запас на начало

периода, приход за период, расход за период, запас на конец периода.

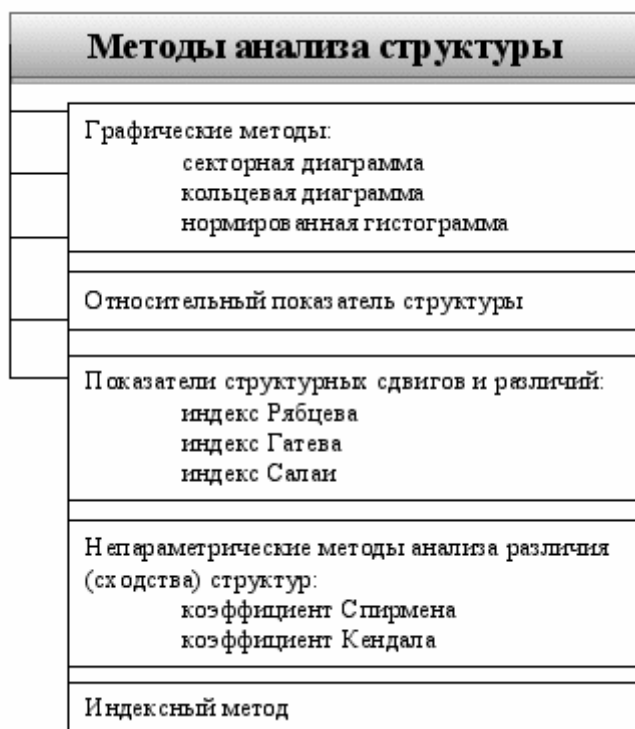
Запас на начало + приход = расход + запас на конец периода.

Для аналитических целей каждая из четырех составляющих делится по различным классификационным признакам на части, группы или подгруппы. Если общий объем признака подразделен по одному группировочному признаку, а затем каждый групповой и общий объемы снова подразделены по другому группировочному признаку, то образуется **многомерная**, в простейшем случае — **двухмерная структура** с пересекающимися признаками.

Двухмерная пересекающаяся структура позволяет рассчитать пять видов структурных показателей (долей). При трех пересекающихся признаках группировки число разных видов структур достигает 19. В общем виде при n взаимопересекающихся признаках структура содержит $(n^3 - n^2 + 1)$ видов долей.

Фасетный метод классификации - это метод, при котором заданное множество делится на группировки независимо, по различным признакам. Она не имеет жесткой структуры и заранее построенных конечных группировок. При ней множество объектов, характеризующихся некоторым набором одинаковых для всех объектов признаков (фасет), может делиться многократно и независимо. В классификаторах фасеты чаще всего располагаются в виде простого перечисления и имеют код. Основное **преимущество** фасетной классификации - гибкость структуры ее построения, так как изменения в любом из фасетов не оказывает влияния на остальные. Фасетный метод классификации позволяет не только образовывать новые классификационные группировки из имеющихся фасетов, но и включать в классификатор без переделки новые и исключать старые фасеты. **Недостатком** фасетной классификации следует считать недостаточно полное использование емкости вследствие отсутствия практически лишних из возможных сочетаний фасетов, непривычность применения, а также сложность использования этого метода для ручной обработки информации. Для цели анализа структуры, а также сравнения двух (или более) структур в динамике используется большое количество статистических методов и подходов которые можно представить в виде

следующей схемы:



Графический сравнительный анализ структуры.

В социально-экономических исследованиях часто возникает ситуации, в которых необходимо анализировать структуры явлений или процессов за ряд периодов. Одним из способов анализа в данном случае является рассмотрение структурных диаграмм. Самой распространенной структурной диаграммой является секторная или круговая

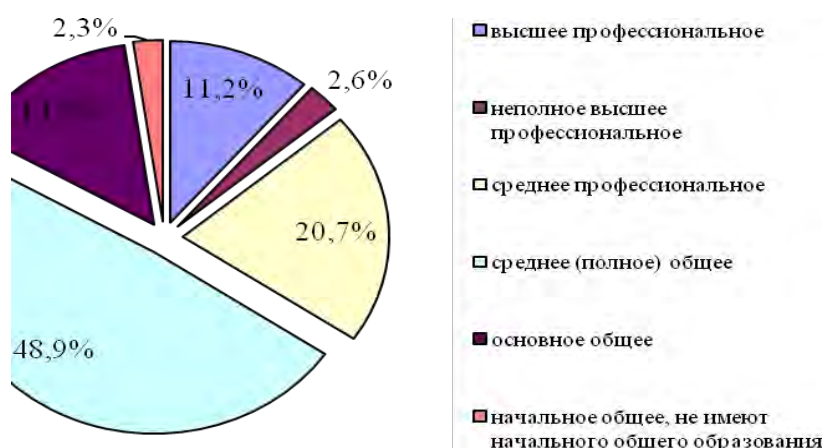


Рисунок - Состав и структура безработных по образованию в 2003г., %

Данный вид диаграмм удобнее всего использовать при иллюстрации структуры явления за один, два или три периода, но на практике может возникнуть ситуация когда необходимо сравнивать структуру за 5 и более периодов. В данном случае необходимо использовать кольцевую диаграмму.

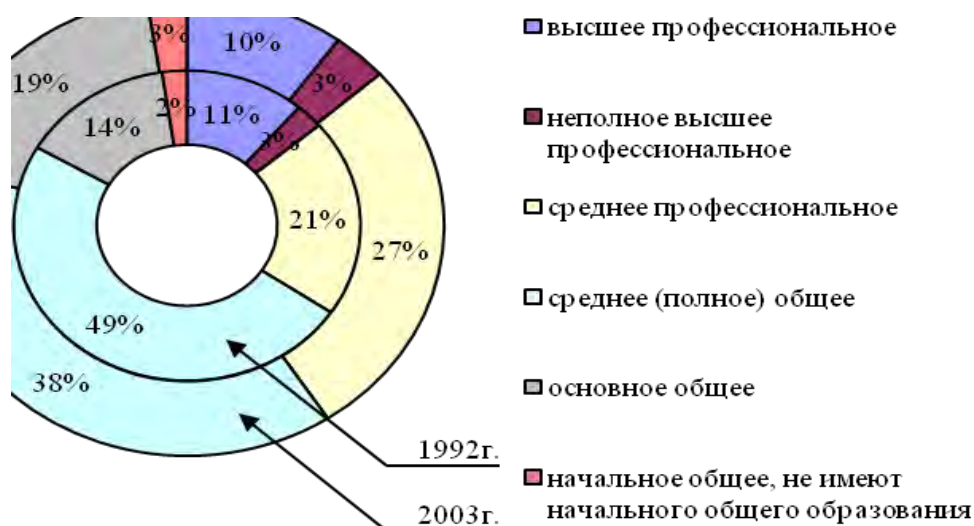


Рисунок - Состав и структура безработных по образованию в 1992г. и 2003г., %

Данный вид структурной диаграммы рекомендуется использовать если необходимо проиллюстрировать данные за 2-3 периода, если периодов более 3 то необходимо использовать особый вид столбиковых диаграмм – нормированную гистограмму.

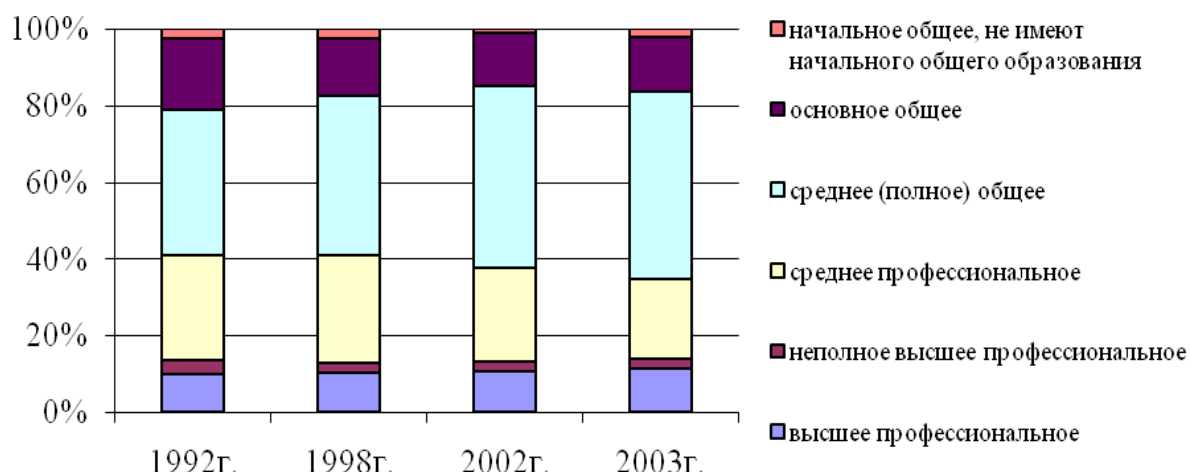


Рисунок - Состав и структура безработных по образованию в 1992г., 1998г., 2002-2003гг., %

С развитием экономической науки и электронно-вычислительной техники все большую роль начинает играть использование математических методов в экономике, или экономико-математическое моделирование.

Любая модель представляет собой упрощенное отображение предмета, процесса или явления. Модель экономического процесса или явления, которая описывает поведение моделируемого объекта с помощью математических выражений, называется экономико-математической моделью.

Такие модели широко используются на практике для анализа и прогнозирования экономической ситуации, а также поддержки процесса принятия решений как на микро-, так и на макроуровне (на отдельных предприятиях и на рынках). Благодаря упрощению модель делает экономическую ситуацию более обозримой, но при этом совокупность данных модели должна содержать информацию, достаточную для того, чтобы судить о главных существенных чертах моделируемого объекта и принимать соответствующие решения.

В процессе производства экономико-математические модели играют вспомогательную роль, окончательное решение является прерогативой руководителя. Это связано со сложностью и недостаточной изученностью комплекса экономики, которая представляет собой сложную систему, отличающуюся целенаправленным поведением своих элементов, интересы которых могут приходить в противоречие с интересами системы в целом.

Моделирование экономических ситуаций средствами исследования операций представляет собой один из видов экономико-математического моделирования. В широком смысле слова под исследованием операций понимают математический аппарат поддержки процесса принятия решения в различных областях человеческой деятельности.

При этом операцией называют любое мероприятие, объединенное общим замыслом и направленное к достижению определенной цели.

Операция обязательно должна быть управляемой, т.е. включать как факторы среды, так и элементы решения.

Чтобы оценить операцию, необходимо определить критерий ее эффективности - показатель того, насколько результат операции соответствует ее целям (в зависимости от предпринятых действий). Чтобы исследовать операцию, строят ее математическую модель - формальные отношения, устанавливающие связь принятого критерия эффективности с объективными условиями и обстоятельствами, определяющими ход операции.

Одним из разделов исследования операций является теория игр (математическая теория конфликтных ситуаций), которая включает в себя теорию принятия решений. Последнюю рассмотрим более подробно.

Элементы теории принятия решений

Условия принятия решений делятся на:

- 1) условия полной определенности;
- 2) условия неопределенности (когда нет полной информации).

Условия неопределенности в свою очередь делятся на:

- а) условия риска (когда известны вероятности поведения природы);
- б) условия “дурной неопределенности” (когда даже эти вероятности неизвестны).

Принятие решений в условиях риска

В случае (а) используются различные средства теории вероятностей. Например, можно подсчитать математические ожидания выигрыша при различных стратегиях A_i (поскольку смешанная стратегия природы известна), и выбрать из них наибольшее.

Например, предприятие может выпускать один из трех видов продукции. Матрица 3 x 4 характеризует прибыль предприятия от выпуска i -го вида продукции (A_i) при j -м состоянии спроса (P_j); $i=1, 3$; $j=1, 4$. Известны вероятности каждого состояния p_j .

P_j	1	2	3	4	\bar{a}_i
A_1	1	4	5	9	5,2
A_2	3	8	4	3	4,5
A_3	4	6	6	2	5,0
p_j	0,1	0,2	0,5	0,2	

$$\bar{a}_i = \sum_{j=1}^4 a_{ij} p_j, \quad i = \overline{1, 3}$$

$$\max_i \{\bar{a}_i\} = 5,2 \quad (i = 1)$$

Следовательно, A_1 – оптимальная стратегия.

Еще раз подчеркнем, что принятое решение является оптимальным не в каждом отдельном случае, а в среднем, при многократном повторении ситуации.

Например, если речь идет о дневной прибыли, то выпуская первый вид продукции в течении одного дня, фирма, возможно, получит и меньшую прибыль, чем получила бы, выпуская другой вид продукции. Прибыль будет оптимальной (максимальной), если просуммировать ее за 100 или 1000 дней, потому что только при большом количестве опытов вступают в действие законы теории вероятностей, но даже в этом случае величина прибыли не будет гарантированной.

Многократное повторение опыта не обязательно должно быть во времени, но может иметь место и в пространстве. Т.е. если прибыль задана для одного филиала фирмы, то речь может идти о применении оптимальной стратегии, например, сотней таких однотипных филиалов.

Другой возможный критерий основан на понятии риска.

Def. Риск – это разность между выигрышем игрока, который он получил бы, если бы знал ситуацию среды (и соответственно выбрал бы стратегию) и выигрышем, который он получит в тех же условиях, используя стратегию A_i : $r_{ij} = \max_k \{a_{kj}\} - a_{ij}$.

Из определения всегда $r_{ij} \geq 0$.

Составим матрицу рисков для предыдущего примера. Для этого вначале в каждом столбце платежной матрицы найдем наибольший элемент (если бы состояние спроса было заранее известно, то была бы выбрана именно эта стратегия). Чтобы подсчитать риск, каждый элемент столбца вычтем из этой величины. Например, при первом состоянии спроса наибольшая прибыль достигается при выпуске третьего вида продукции, и равна 4. Если использовать первую стратегию, то прибыль будет равна 1. Следовательно, риск $r_{11} = 4 - 1 = 3$ (именно столько мы рискуем недополучить, используя первую стратегию в первой ситуации по сравнению с максимально возможным выигрышем). Аналогично $r_{21} = 4 - 3 = 1$; $r_{31} = 4 - 4 = 0$. При втором состоянии спроса наибольшая прибыль – 8 - может быть получена при выпуске второго вида продукции, поэтому все элементы второго столбца вычитаются из 8, и т.д.

Найдем с помощью матрицы рисков оптимальную стратегию (для которой ожидаемый риск – наименьший).

P_j	1	2	3	4	\bar{r}_i
A_1	3	4	1	0	1,6
A_2	1	0	2	6	2,3
A_3	0	2	0	7	1,8
p_j	0,1	0,2	0,5	0,2	

$$\bar{r}_i = \sum_{j=1}^4 r_{ij} p_j, \quad i = \overline{1, 3}$$

$$\min_i \{\bar{r}_i\} = 1,6 \quad (i = 1)$$

Следовательно, A_1 – оптимальная стратегия; т.е. результат совпал с предыдущим.

Покажем, что решения, найденные этими двумя способами, всегда будут совпадать.

$$\bar{a}_i + \bar{r}_i = \sum_j a_{ij} p_j + \sum_j r_{ij} p_j = \sum_j p_j (a_{ij} + \max_i \{a_{ij}\} - a_{ij}) = \sum_j p_j \max_i \{a_{ij}\} = \text{const}$$

Итак, эта сумма является постоянной (взвешенное среднее максимальных элементов по столбцам платежной матрицы).

$$\overline{a_i} + \overline{r_i} = \text{const} \Rightarrow \overline{r_i} = \text{const} - \overline{a_i}$$

Эта величина обращается в минимум тогда же, когда $\overline{a_i}$ – в максимум.

Могут быть использованы и другие критерии.

Отметим, что в условиях риска всегда будет получено решение в чистых стратегиях.

Принятие решений в условиях “дурной” неопределенности

При принятии решений в условиях “дурной” неопределенности можно выделить два основных подхода. Первый из них – попытаться свести ситуацию к условиям риска, т.е. некоторым способом определить вероятности возникновения различных ситуаций среды.

Пример такого подхода – использование критерия Лапласа. Он основан на принципе недостаточного обоснования: если нет оснований считать, что вероятности состояний различны, их можно считать равными, т.е.

$$p_1 = \dots = p_j = \dots = p_n = 1/n.$$

Изменим условия предыдущего примера. Пусть теперь вероятности возникновения того или иного состояния спроса неизвестны. Применяя критерий Лапласа, будем считать их равными. Так как всего состояний 4, вероятность каждого из них будет 1/4. Тогда ожидаемый выигрыш при использовании каждой стратегии можно определить, как простое среднее элементов по строкам матрицы:

π_j	1	2	3	4	$\overline{a_i}$
A_1	1	4	5	9	4,75
A_2	3	8	4	3	4,5
A_3	4	6	6	2	4,5
p_j	0,25	0,25	0,25	0,25	

$$\overline{a_i} = \left(\sum_{j=1}^4 a_{ij} \right) / 4, \quad i = \overline{1, 3}$$

$$\max_i \{ \overline{a_i} \} = 4,75 \quad (i = 1)$$

Следовательно, A_1 – оптимальная стратегия, и ничего не зная о вероятностях состояний спроса, все равно следует выпускать первый вид продукции.

Другой способ в рамках того же подхода можно применить, если вероятности состояний можно оценить в ранговой шкале (т.е. упорядочить их, начиная с наиболее вероятных). Тогда предполагают, что эти вероятности пропорциональны членам убывающей арифметической прогрессии:

$$p_1:p_2:p_3:\dots:p_j:\dots:p_n = n:(n-1):(n-2):\dots:(n-j+1):\dots:1.$$

Так как должно быть $\sum_j p_j = 1$, а $\sum_{j=1}^n (n-j+1) = \sum_{j=1}^n (j) = \frac{n(n+1)}{2}$, то $p_j = \frac{2(n-j+1)}{n(n+1)}$, $j = \overline{1, n}$ (соответствующий член прогрессии отнесен к сумме прогрессии).

Пусть теперь в условиях того же примера известно, что $p_3 > p_4 > p_2 > p_1$. Тогда предположим, что $p_3:p_4:p_2:p_1 = 4:3:2:1$. Здесь $2/(n(n+1)) = 2/(4*5) = 0,1$; вероятности будут равны 0,4; 0,3; 0,2 и 0,1. Ожидаемый выигрыш тогда будет подсчитан следующим образом:

p_j	1	2	3	4	$\overline{a_i}$
A_1	1	4	5	9	5,6
A_2	3	8	4	3	4,5
A_3	4	6	6	2	5
p_j	0,1	0,2	0,4	0,3	

$$\max_i \{\overline{a_i}\} = 5,6 (i = 1)$$

Здесь оптимальная стратегия – снова A_1 .

Другой возможный способ в рамках того же подхода – использование экспертных оценок для определения вероятностей.

Принципиально другой подход – не сводить условия “дурной” неопределенности к условиям риска. Рассмотрим некоторые критерии, который используются в этом случае.

1. Критерий Вальда – максиминный выигрыш:

$$W = \max_i \min_j \{a_{ij}\}.$$

Это критерий крайнего пессимизма, критерий осторожности, так как природа здесь считается противником, выбирающим наихудшую стратегию для лица, принимающего решение.

2. Критерий Сэвиджа – минимаксный риск:

$$S = \min_i \max_j \{r_{ij}\}.$$

Этот критерий также пессимистический, но здесь худшим считается не меньший выигрыш, а большая потеря выигрыша по сравнению с возможным в данных условиях.

3. Критерий пессимизма-оптимизма Гурвица:

$$H = \max_i \left\{ h \min_j \{a_{ij}\} + (1-h) \max_j \{a_{ij}\} \right\}, \quad 0 \leq h \leq 1.$$

При $h=1$ этот критерий совпадает с пессимистическим критерием Вальда.

При $h=0$ он превращается в критерий оптимизма $\max_i \max_j \{a_{ij}\}$.

Таким образом, коэффициент h выражает меру пессимизма лица, принимающего решение и выбирается из субъективных соображений (чем опаснее ситуация, тем ближе он должен быть к 1).

Критерий Гурвица можно построить и для рисков:

$$H = \min_i \left\{ h \max_j \{r_{ij}\} + (1-h) \min_j \{r_{ij}\} \right\}.$$

Применим эти критерии к тому же примеру, в котором теперь вероятности возникновения того или иного состояния спроса неизвестны (оптимальную стратегию обозначим A^*):

$$a_{ij} = \begin{pmatrix} 1 & 4 & 5 & 9 \\ 3 & 8 & 4 & 3 \\ 4 & 6 & 6 & 2 \end{pmatrix} \quad \begin{matrix} \min_j \{a_{ij}\} \\ 1 \\ 3 \\ 2 \end{matrix} \quad W = 3 \quad A^* = A_2$$

$$r_{ij} = \begin{pmatrix} 3 & 4 & 1 & 0 \\ 1 & 0 & 2 & 6 \\ 0 & 2 & 0 & 7 \end{pmatrix} \quad \begin{matrix} \max_j \{r_{ij}\} \\ 4 \\ 6 \\ 7 \end{matrix} \quad S = 4 \quad A^* = A_1$$

Подсчитаем критерий Гурвица для различных h :

$a_{ij} = \begin{pmatrix} 1 & 4 & 5 & 9 \\ 3 & 8 & 4 & 3 \\ 4 & 6 & 6 & 2 \end{pmatrix}$	$\min_j \{a_{ij}\}$	$\max_j \{a_{ij}\}$	H	$h = 0,8$	$h = 0,5$	$h = 0,2$
	1	9	$h + 9(1-h) = 9 - 8h$	3,6	5	8,4
	3	8	$3h + 8(1-h) = 8 - 5h$	4	5,5	7
	2	6	$2h + 6(1-h) = 6 - 4h$	2,8	4	5,2

При $h=0,8$ (наиболее пессимистический из рассмотренных вариантов) и при $h=0,5$ $A^*=A_2$. В более оптимистическом варианте ($h=0,2$) $A^*=A_1$.

Применять к данному случаю критерий Гурвица для рисков не имеет смысла, так как в матрице рисков здесь во всех строках имеются нули (минимальные риски по строкам, которые следует умножить на $(1-h)$, везде будут равны 0).

Итак, в условиях “дурной” неопределенности с точки зрения уменьшения риска целесообразно выпускать первый вид продукции, а с точки зрения получения наибольшей прибыли – второй. Выбор первого вида продукции с точки зрения получения прибыли предполагает значительный оптимизм.

Проиллюстрируем применение рассмотренных методов следующей схемой:

Методы принятия решения

а) в условиях риска:

- 1) расчет
ожидаемого
выигрыша;
- 2) расчет
ожидаемого риска;
и т.п.

б) в условиях «дурной неопределенности»

- | | |
|---|---|
| <p>путем сведения
ситуации к условиям
риска (см. (а)),
вероятности состояний
определяются:</p> <ul style="list-style-type: none"> 1) по критерию
Лапласа; 2) через ранговые
оценки вероятностей; 3) экспертным путем | <p>не сводя ситуацию к
условиям риска, по
критериям:</p> <ul style="list-style-type: none"> 1) Вальда; 2) Сэвиджа; 3) Гурвица для
выигрышей или
рисков |
|---|---|

Постановка задачи линейного программирования. Постановка задачи производственного планирования

Построим математическую модель следующей экономической ситуации.

Кондитерская фабрика при производстве двух видов карамели «Снежинка» и «Яблочная» - использует три вида основного сырья: сахарный песок, патоку и фруктовое пюре. Запасы сырья составляют соответственно 800 т, 600 т и 120 т. Прибыль от реализации 1 т «Снежинки» составляет 108 ден.ед., а «Яблочной» - 140 ден.ед. На выпуск 1 т «Снежинки» расходуется 0,8 т сахара, 0,2 т патоки и 0,01 т фруктового пюре, а на выпуск 1 т «Яблочной» - соответственно по 0,5 т, 0,4 т и 0,1 т этих видов сырья.

Необходимо найти план производства карамели, который даст наибольшую прибыль.

Описанная ситуация представляет собой задачу производственного планирования (или задачу об использовании ресурсов). Чтобы построить математическую модель данной ситуации, прежде всего введем переменные, значения которых определяют принимаемое решение. В самом деле, что в данном случае означает «план производства»? Очевидно, следует найти, сколько именно нужно выпускать продукции каждого вида.

Обозначим x_1 - производство карамели «Снежинка», т; x_2 - производство карамели «Яблочная», т.

Теперь задумаемся о цели данной операции. В данном случае целью является получение как можно более высокой общей выручки или прибыли. Каким образом можно вычислить ее через исходные данные задачи и введенные нами обозначения?

Прибыль от производства карамели «Снежинка» составит $108x_1$ ден.ед. В самом деле, если выпустить 10 т этой карамели ($x_1=10$), то прибыль от нее составит $108 \cdot 10 = 1080$ (ден.ед.), если выпустить 2 т, то $108 \cdot 2 = 216$ (ден.ед.); если вообще не выпускать эту карамель, то и прибыли от нее не будет – $108 \cdot 0 = 0$, и т.д. Аналогично прибыль от «Яблочной» составит $140x_2$ ден.ед.

Общая прибыль представляет собой сумму прибыли от двух видов карамели. Необходимо максимизировать прибыль от всей карамели, т.е. $108x_1 + 140x_2$.

Очевидно, что если бы значения переменных могли быть любыми, выручка (или прибыль) могла бы возрасти до бесконечности. Однако это не так, на выпуск продукции наложены некоторые ограничения. Прежде всего, они связаны с ограниченными запасами ресурсов.

На 1 т карамели «Снежинка» затрачивается 0,8 т сахарного песка. Следовательно, на x_1 т этой карамели будет затрачено $0,8x_1$ т (т.е. 1,6 т, если выпускается 2 т такой карамели, 80 т, если выпускается 100 т такой карамели, и т.д.). Аналогично на производство карамели «Яблочная» будет затрачено $0,5x_2$ т сахарного песка. Таким образом, общие затраты сахарного песка составят $0,8x_1 + 0,5x_2$ т. Поскольку его запас составляет 800 т, можно записать: $0,8x_1 + 0,5x_2 \leq 800$.

Аналогично для патоки следует построить ограничение $0,2x_1 + 0,4x_2 \leq 600$, а для фруктового пюре $0,01x_1 + 0,1x_2 \leq 120$. Таким образом, построено три ограничения.

По смыслу задачи обе переменные неотрицательны: $x_1 \geq 0$, $x_2 \geq 0$.

Итак, математическую модель можно построить следующим образом:

$$\begin{aligned} & \max 108x_1 + 140x_2 \\ & \begin{cases} 0,8x_1 + 0,5x_2 \leq 800 \\ 0,2x_1 + 0,4x_2 \leq 600 \\ 0,01x_1 + 0,1x_2 \leq 120 \\ x_{1,2} \geq 0 \end{cases} \end{aligned}$$

Все выражения, входящие в запись данной математической модели, представляют собой линейные функции. Построенная модель представляет собой задачу линейного программирования.

Основные понятия линейного программирования

Задача линейного программирования в общем виде может быть записана следующим образом:

$$\begin{cases} \max \\ \min \end{cases} \sum_{j=1}^n c_j x_j \quad (1)$$

$$\begin{cases} \sum_{j=1}^n a_{ij} x_j & R_i & b_i, & i = \overline{1, m} \end{cases} \quad (2)$$

где $R_i \in \{\leq; \geq; =\}$;

x_j – переменные;

a_{ij}, b_i, c_j – константы;

n – число переменных;

m – число ограничений.

Распространенным является случай, когда на все переменные накладываются ограничения неотрицательности, т.е. $x_j \geq 0, j = \overline{1, n}$. При этом,

говоря о размерности задачи (числе переменных и ограничений), ограничения на знак переменных обычно не считают.

Выражение (1) представляет собой целевую функцию задачи линейного программирования (или функцию цели), которую необходимо максимизировать или минимизировать. Например, для задачи, поставленной в предыдущем разделе, целевой функцией является выражение $(108x_1 + 140x_2)$ – именно оно отражает цель операции. Выражения (2) представляют собой систему ограничений, которая может состоять из уравнений и нестрогих неравенств. В рассмотренной задаче производственного планирования такая система состояла только из неравенств: трех ограничений по запасам ресурсов и двух ограничений неотрицательности переменных.

Любой набор значений переменных, т.е. вектор чисел $X = (x_1, x_2, \dots, x_n)$, называется планом задачи линейного программирования. Например, для той же задачи любые два числа будут представлять собой план. План (10; 20) означает, что $x_1=10$, а $x_2=20$, т.е. выпуск карамели «Снежинка» составит 10 т, а выпуск карамели «Яблочная» - 20 т. План (2000; 1000) означает, что выпуск карамели «Снежинка» составит 2000 т, а выпуск карамели «Яблочная» - 1000 т. План (0; 0) означает, что карамель не будет выпускаться вообще. План (0; -10) означает, что карамель «Снежинка» не выпускается, а выпуск карамели «Яблочная» составит -10 т. Последнее, вообще говоря, очевидно бессмысленно. И т.д. Если в задаче 5 переменных, то любые 5 чисел будут представлять собой план, если 10 переменных, - то любые 10 чисел; и т.д.

План, удовлетворяющий системе ограничений, называется допустимым. Чтобы проверить, является ли план допустимым или нет, необходимо подставить его в систему ограничений. Если при этом все

уравнения и неравенства истинны, то план допустимый. Если хотя бы одно ложно, то план – недопустимый.

Проверим на допустимость план (10; 20):

$$\begin{cases} 0,8x_1 + 0,5x_2 = 0,8*10 + 0,5*20 = 18 \leq 800 \\ 0,2x_1 + 0,4x_2 = 0,2*10 + 0,4*20 = 10 \leq 600 \\ 0,01x_1 + 0,1x_2 = 0,01*10 + 0,1*20 = 2,1 \leq 120 \\ x_1 = 10 \geq 0 \\ x_2 = 20 \geq 0 \end{cases}$$

Следовательно, этот план допустимый. Действительно, кондитерская фабрика может выпускать карамель таким образом. При этом будет затрачено всего 18 т сахарного песка, 10 т патоки и 2,1 т фруктового пюре, а в запасе имеется значительно больше.

Если проверить на допустимость план (0; 0), в результате будет получена система неравенств:

$$\begin{cases} 0 \leq 800 \\ 0 \leq 600 \\ 0 \leq 120 \\ 0 \geq 0 \\ 0 \geq 0 \end{cases}$$

Все эти неравенства истинны. Следовательно, и этот план допустимый. Впрочем, и без того очевидно, что кондитерская фабрика может не выпускать продукцию.

Проверим на допустимость план (2000; 1000):

$$\begin{cases} 0,8x_1 + 0,5x_2 = 0,8*2000 + 0,5*1000 = 2100 \leq 800 \\ 0,2x_1 + 0,4x_2 = 0,2*2000 + 0,4*1000 = 800 \leq 600 \\ 0,01x_1 + 0,1x_2 = 0,01*2000 + 0,1*1000 = 120 \leq 120 \\ x_1 = 2000 \geq 0 \\ x_2 = 1000 \geq 0 \end{cases}$$

Первые два утверждения – ложные. На самом деле, $2100 > 120$, а $800 > 600$. Следовательно, этот план недопустимый (вообще говоря, убедившись, что первое неравенство – ложное, можно было выполнение остальных и не проверять). Итак, мы убедились, что фабрика не может выпускать карамель таким образом, так как ей не хватит для этого сахарного песка и патоки.

Если подставить в систему ограничений план $(0; -10)$, то все они выполняются, кроме последнего ограничения: $-10 < 0$. Следовательно, раз хотя бы одно ограничение не выполняется, и этот план недопустимый (выпустить -10 т карамели невозможно).

Вся совокупность допустимых планов представляет собой область допустимых планов (ОДП) задачи.

Допустимый план, на котором достигается максимальное или минимальное значение целевой функции, называется оптимальным планом, а само это значение – оптимумом задачи.

Для рассмотренной задачи оптимальным планом будет такой допустимый план производства карамели, который позволит получить наибольшую прибыль. Сама эта величина прибыли – самая высокая, какую только можно получить, – будет представлять собой оптимум.

Следует отметить, что оптимальных планов в задаче может быть более одного, но значение целевой функции на них должно быть одинаковым, т.е. оптимум всегда один. Например, в задаче производственного планирования иногда можно найти разные способы произвести продукцию таким образом, чтобы прибыль (или выручка) были максимальными. Но все они дадут одну и ту же – максимальную – величину прибыли.

Таким образом, с точки зрения изучаемой дисциплины употребление таких словосочетаний, как «самый оптимальный» или «наиболее оптимальный», является неграмотным. Слово «оптимальный» не сочетается со словами «самый», «наиболее», «очень» и т.п., как любое прилагательное в превосходной степени (нельзя сказать «самый наилучший»). На самом деле, такое выражение просто бессмысленно – ведь если некоторый план действий – «самый оптимальный», значит, есть и другие оптимальные планы, «менее оптимальные», чем он. Но раз они чем-то хуже, значит, они оптимальными не являются.

Можно доказать, что если оптимальный план существует, то он либо единственный, либо их бесконечно много^{*}; любая смесь (взвешенная сумма) оптимальных планов является оптимальным планом.

Решить задачу линейного программирования означает найти все ее оптимальные планы и оптимум, хотя часто ограничиваются нахождением одного из возможных решений. Если задача линейного программирования не имеет решений, необходимо установить причину ее неразрешимости. Это может быть одна из двух причин:

- ее ОДП пуста (т.е. система ограничений несовместна),
- целевая функция не ограничена на ОДП (т.е. ее значение можно увеличивать или уменьшать до бесконечности).

Задачи линейного программирования, как экономико-математические модели, находят очень широкое применение. Рассмотренная выше задача производственного планирования представляет собой лишь одну из возможных экономических интерпретаций задачи линейного программирования, наиболее традиционную. Множество других

^{*} За исключением задач, в которых переменные могут принимать только целые значения.

экономических ситуаций может быть описано в тех же математических терминах, что делает возможным применение к их решению одного и того же математического аппарата.

Для того чтобы построить математическую модель экономической ситуации в виде задачи линейного программирования, прежде всего необходимо ввести переменные задачи. Они должны быть введены таким образом, чтобы их значения определяли принимаемое решение (получив значения переменных, мы получаем ответ на поставленный вопрос).

Затем определяют цель, критерий эффективности операции, ту величину, которую необходимо экстремизировать в задаче. Ее выражают через введенные переменные - получают линейное выражение для целевой функции.

После этого необходимо установить, чем ограничивается рост или уменьшение целевой функции, т.е. определить ограничения задачи. Их нужно также выразить через переменные и записать в виде системы уравнений и неравенств.

Кроме того, при построении модели полезно воспользоваться следующими рекомендациями. При определении переменных следует заранее обдумать, позволят ли они отразить в модели все условия задачи (если известно, что нет избыточных условий). В конкретной задаче указывают единицы измерения для переменных. Если в исходных данных задачи одна и та же величина измеряется в различных единицах (например, масса в граммах, килограммах, тоннах), то необходимо перевести эти данные в одни и те же единицы измерения. Выражая целевую функцию и ограничения через переменные, следует проверить, какими единицами будут измеряться полученные величины и не являются ли они бессмысленными с экономической точки зрения (например, не измеряются ли левая и правая части ограничений в разных единицах). Отдельно следует обдумать ограничения на знак переменных.

В некоторых задачах переменные могут принимать только целые значения. Этот факт также необходимо записать в виде ограничения: $X \in Z$. Такое ограничение выводит поставленную задачу из класса задач линейного программирования в класс задач целочисленного линейного программирования. Однако, рассмотренных здесь понятий достаточно для того, чтобы построить математическую модель и для целочисленной задачи.

Приведем некоторые примеры экономических задач, математические модели которых можно построить в виде задач линейного программирования. Следует отметить, что здесь будут рассмотрены далеко не все такие задачи; а кроме того, даже те, что рассмотрены, существуют в разнообразных модификациях, и классификация этих задач в большой мере условна, различается у разных авторов.

Студентам предлагается самостоятельно изучить примеры постановки задачи о диете (о составлении рациона, о смеси), задачи о раскрое материалов, задачи о загрузке транспорта и др.

Формы записи задачи линейного программирования

Формулы (1) и (2) описывают задачу линейного программирования в смешанной форме. Кроме того, существует несколько специальных форм записи таких задач.

Задачу с неотрицательными переменными, все остальные ограничения которой имеют форму уравнений, будем называть канонической формой записи задачи линейного программирования:

Любая задача в смешанной форме может быть приведена к канонической. Это делается путем введения неотрицательных дополнительных переменных, которые прибавляются к левой части неравенства либо вычитаются из нее (после этого знак неравенства заменяют

на «равно»). Дополнительные переменные представляют собой разность между частями неравенства.

Неотрицательности переменных добиваются заменой неограниченных по знаку переменных на разность двух неотрицательных. Подробнее этот материал рекомендуется изучить самостоятельно.

Задачу с неотрицательными переменными, все остальные ограничения которой имеют форму неравенств одного знака (\leq , если задача на максимум, и \geq , если на минимум), будем называть стандартной формой записи задачи линейного программирования. Любая задача в смешанной форме может быть приведена к любой из стандартных форм. Подробнее этот материал рекомендуется изучить самостоятельно.

Если задача линейного программирования приведена к стандартной или канонической форме, ее удобно кратко записать в матричной форме.

Спецификация модели

Эконометрика – это наука, изучающая количественные взаимосвязи экономических показателей на основе использования аппарата теории вероятностей и математической статистики.

Эта наука возникла в результате взаимодействия и объединения экономической теории, статистики (как экономической, так и математической) и других разделов математики. Методы эконометрики позволяют выявлять новые, ранее не известные связи между экономическими показателями, а также уточнять или отвергать гипотезы о существовании таких связей.

Эконометрическая модель, таким образом, представляет собой формализованную математическую модель взаимосвязи между экономическими показателями, построенную на основе использования статистических методов.

Можно сказать, что она представляет собой частный случай экономико-математической модели, т.е. экономическую модель, представленную в математической форме.

Уравнение регрессии. Взаимосвязи между показателями, отражаемые в эконометрической модели, обычно описываются **уравнением регрессии**. Приведем упрощенный пример. Предположим, что в начальный момент времени цена на некоторую продукцию составляет 5 ден.ед., при этом каждый месяц она возрастает на 10%, т.е. в 1,1 раза. Тогда уравнение регрессии, отражающее взаимосвязь между двумя показателями: ценой продукции и временем, будет выглядеть следующим образом:

$$y_t = 5 * 1,1^t + \varepsilon_t,$$

где t – время, мес.,

y_t – цена в момент времени t ,

ε_t - случайная компонента процесса на момент времени t .

Первый вопрос, который решают при построении эконометрической модели, - это вопрос спецификации.

Спецификация модели – это математическая форма записи уравнения зависимости результирующей переменной от одного или нескольких факторов. По сути, она представляет собой отбор факторов, включаемых в модель, и выбор вида уравнения регрессии.

В зависимости от спецификации эконометрических моделей классифицируют на несколько типов. Соответственно, основных классификационных признаков, связанных с классификацией, два – количество включенных факторов (т.е. экономических показателей) и тип зависимости между показателями в уравнении регрессии. Кроме того, эконометрические модели классифицируют по типу исходных данных.

По количеству включенных факторов различают два типа эконометрических моделей: модели парной и множественной регрессии.

Парная регрессия (ее еще называют – простая) представляет собой регрессию между двумя показателями, двумя переменными.

Приведенный выше пример относится к парной регрессии. В общем виде уравнение парной регрессии можно записать следующим образом:

$$y = f(x) + \varepsilon,$$

где y – зависимая переменная (результативный признак);

x – независимая переменная (признак-фактор);

$f(x)$ – функция, отражающая регрессионную зависимость;

ε - случайная компонента, которая характеризует отклонение реального значения результативного признака от теоретического.

Множественная регрессия представляет собой зависимость между результативным признаком и двумя либо большим числом факторов.

В общем виде ее можно записать следующим образом:

$$y = f(x_1, x_2, \dots, x_n) + \varepsilon,$$

где y – результативный признак;

$n \geq 2$ – число независимых факторов, включенных в модель;

x_1, x_2, \dots, x_n – признаки-факторы;

$f(X)$ – функция, отражающая регрессионную зависимость;

ε – случайная компонента.

При использовании моделей множественной регрессии обычно, помимо построения самого уравнения регрессии и определения совокупного влияния факторов на моделируемый показатель, определяют также влияние на результат каждого фактора в отдельности.

По-другому признаки-факторы еще называют объясняющими, предикторными, экзогенными переменными, регрессорами. Зависимую, результирующую переменную y иногда называют эндогенной.

По типам зависимости, используемым в модели, можно выделить следующие типы эконометрических моделей:

1) линейная (если используется линейная зависимость между показателями) ($y = ax + b$ или $y = a_1x_1 + a_2x_2 + \dots + a_nx_n + b$; парную линейную регрессию еще иногда называют простой регрессией);

2) нелинейная (если используется нелинейная связь). Здесь, в свою очередь, могут быть использованы различные функции:

а) полиномы различных степеней, начиная со второй, т.е.

- квадратическая функция ($y = ax^2 + bx + c$),

- кубическая функция ($y = ax^3 + bx^2 + cx + d$)

- и т.д.

б) степенная функция ($y = ax^b$)

в) гиперболическая функция ($y = a/x + b$)

Формулы для нелинейных функций здесь приведены для случая, когда имеется один признак-фактор, хотя эти функции можно использовать и в случае множественной регрессии.

г) показательная (экспоненциальная) функция ($y = ae^{bx}$). Можно показать, что показательная и экспоненциальная функция – одно и то же. Действительно, пусть $y = a\beta^x = a(e^{\ln \beta})^x = ae^{x \ln \beta} = ae^{bx}$, где $b = \ln \beta$.

д) другие нелинейные функции (простая модифицированная экспонента ($y = k - ae^{bx}$); логистическая кривая ($y = k / (1 - ae^{bx})$); функция Гомпертца $y = ka^{(bx)}$ и т.д.)

Классификация эконометрических моделей по типам данных

Кроме того, эконометрические модели можно классифицировать по типам данных. А именно, исходные данные для модели могут быть пространственными либо временными рядами.

Пространственные данные представляют собой данные об экономических показателях на определенный момент времени (например, данные о среднем уровне дохода и среднем уровне спроса на некоторую продукцию в различных регионах позволяют исследовать зависимость между спросом и доходом).

Временные ряды позволяют исследовать зависимость показателя от времени (например, по данным о спросе на продукцию в одном и том же регионе в разные моменты времени позволяют провести такое исследование для спроса).

Основные проблемы, возникающие в процессе моделирования, и связанные с ними возможные ошибки. Выделим основные проблемы, возникающие в процессе моделирования, и связанные с ними возможные ошибки:

1) проблема отбора факторов, включаемых в модель

а) количество факторов

б) качественный состав факторов

Эти две стороны одной проблемы тесно взаимосвязаны. Содержательный отбор факторов очень важен, поскольку он лежит в основе спецификации модели. Чем больше факторов, влияющих на результат, будет учтено, тем более адекватной, близкой к реальности будет построенная модель. Но, с другой стороны, существует опасность включить в модель лишние факторы, которые не являются существенными в данном исследовании. Это может привести к очевидно нелепым результатам (в качестве примера можно привести анекдот о вреде огурцов: «Сколько людей, умерших в раннем возрасте, употребляло огурцы? Почти 100%, - следовательно, огурцы вредны для здоровья»).

Кроме того, чем больше факторов включено в модель, тем она сложнее, и выигрыш в адекватности может быть иногда сведен на нет трудностями, связанными с построением и использованием модели.

Иногда количество факторов, включаемых в модель, можно уменьшить за счет их агрегирования, т.е. объединения нескольких факторов в один. Кроме того, включив в модель время, можно иногда учесть не учтенные в явном виде факторы, связанные со временем.

Ошибки в спецификации модели могут быть также допущены за счет того, что признаки-факторы влияют не только на результат, но и друг на друга.

2) проблема определения связи между факторами

а) выбор функции для моделирования

б) определение параметров этой функции

Наиболее простой функцией является линейная, и в случае парной регрессии для нее необходимо определить всего два параметра. Тем не менее, не все зависимости можно моделировать с помощью этой функции. При использовании нелинейных функций также необходимо помнить о том, что выигрыш в адекватности сопровождается ростом сложности модели.

3) определение исходных данных, на основании которых будет строиться модель

а) объема выборки

б) состава выборки

При построении эконометрической модели невозможно учесть все возможные значения показателей, так как обычно их очень много и даже иногда бесконечно много. Например, исследуя зависимость спроса от дохода путем опроса потребителей, обычно невозможно опросить всех потенциальных покупателей, и опрашивают только некоторых из них. В таком исследовании все покупатели представляют собой генеральную совокупность, а опрошенные (учтенные в модели) – выборку.

Если выборка мала, то имеется мало возможных комбинаций значений показателей, и поэтому велика вероятность случайного обнаружения комбинации значений, показывающих сильную зависимость, которой на самом деле нет. Например, опросив всего пару человек из сотен, мы можем случайно столкнуться с людьми, которые при высоком доходе предъявляют небольшой спрос на данный товар, из чего сделаем неправильный вывод о зависимости между этими показателями.

Другой пример, который часто используется в теории вероятности, - подбрасывание монеты (здесь генеральная совокупность - бесконечно большое число опытов). В нормальных условиях герб или решка выпадают примерно с одинаковой частотой. Предположим, что монета повреждена, и герб выпадает в 60% случаев. Пусть объем выборки – 100 отдельных экспериментов, и герб выпал 58 раз. Это с очень большой неуверенностью позволит предположить, что монета неправильно сбалансирована, поскольку такое отклонение от ожидаемых 50% не столь велико. Для подтверждения предположения о повреждении монеты следовало бы провести большее количество экспериментов, например, 1000. Однако, если бы монета была повреждена более существенно (например, герб выпадал бы в 99% случаев),

и в выборке из 100 экспериментов он выпал бы 98 раз, то тогда это послужило бы явным признаком повреждения монеты.

Минимальный объем выборки. Из последнего примера видно, что необходимый минимальный объем выборки зависит от того, насколько сильную связь исследуют.

Строя эконометрическую модель на основе выборки заданного объема, мы выдвигаем гипотезу, что существует достаточно сильная связь между показателями. В реальности эта связь может отсутствовать (назовем это **нулевой гипотезой**). Насколько наша гипотеза является значимой, насколько мы можем быть в ней уверены? Она тем более значима, чем меньше вероятность ее принять в случае, если на самом деле реализовалась нулевая гипотеза. Т.е. вероятность того, что мы установили наличие связи, когда ее на самом деле нет, должна быть мала. Эта вероятность функционально зависит от объема выборки и от того, насколько сильную зависимость мы хотим выявить. Для определения этой вероятности используют статистические таблицы, в которых отражены значения соответствующих функций. В свою очередь, при определении объема выборки для построения эконометрической модели необходимо заранее задаться этой вероятностью.

Кроме того, необходимо помнить, что число наблюдений в выборке должно превосходить число признаков в несколько раз, чтобы параметры уравнения множественной регрессии были статистически надежными.

Большую роль также играет качественный состав выборки. Она обязательно должна быть репрезентативной, т.е. правильно отражать свойства генеральной совокупности. Для этого при определении методики отбора исходных данных необходимо исключить факторы, которые сами по себе могут повлиять на результат. Например, проводя телефонный опрос потребителей с целью выяснить их отношение к определенному виду продукции, можно получить данные в основном о тех респондентах, которые

часто находятся дома, у телефона. Это может само по себе влиять на их потребительские предпочтения и, таким образом, исказить реальную ситуацию.

Коэффициент корреляции

При проведении эконометрического исследования, как правило, помимо построения уравнения регрессии, рассчитывают также показатели тесноты связи между параметрами. Один из таких показателей – **коэффициент корреляции**. Он измеряет тесноту линейной связи между переменными.

Коэффициент корреляции между переменными x и y рассчитывается по формуле:

$$r_{xy} = \frac{\text{Cov}(x, y)}{\sigma(x)\sigma(y)},$$

где $\text{Cov}(x, y) = \frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y}) = \overline{xy} - \bar{x} * \bar{y}$ - коэффициент

ковариации между этими переменными;

m – число наблюдений;

$x_1, x_2, \dots, x_m; y_1, y_2, \dots, y_m$ – значения переменных;

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i, \quad \bar{y} = \frac{1}{m} \sum_{i=1}^m y_i, \quad \overline{xy} = \frac{1}{m} \sum_{i=1}^m x_i y_i - \text{средние значения } x, y \text{ и } xy;$$

$$\sigma(x) = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2}, \quad \sigma(y) = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \bar{y})^2} - \text{среднеквадратические}$$

(стандартные) отклонения значений переменных (СКО).

При этом $\text{Cov}(x, x) = \sigma^2(x)$ а $\text{Cov}(y, y) = \sigma^2(y)$, т.е. ковариация переменной с самой собой равна ее дисперсии. Поэтому корреляция признака с самим собой всегда равна единице.

Коэффициент корреляции может принимать значения в диапазоне от -1 (отрицательная корреляция) до 1 (положительная корреляция). Чем ближе он по модулю к единице, тем теснее линейная связь. Если модуль равен 1, то связь функциональная (т.е. нет случайных отклонений). Если его значение приближается к нулю, линейная связь между переменными отсутствует. Отметим, что коэффициент корреляции может равняться нулю тогда и только

тогда, когда ковариация признаков равна нулю (это следует из формулы (1.4)): $r_{xy} = 0 \Leftrightarrow \text{Cov}(x, y) = 0$.

Поскольку в числителе и знаменателе формулы находятся величины, измеряемые в одних и тех же единицах измерения (перемножаются единицы измерения показателей), коэффициент корреляции является величиной безразмерной.

Существуют и другие формулы для расчета коэффициента корреляции, которые дают тот же результат.

Для измерения тесноты связи используется также коэффициент детерминации, который представляет собой квадрат коэффициента корреляции и будет впоследствии рассмотрен более подробно. Отметим только, что этот коэффициент изменяется от 0 до 1, и чем он ближе к единице, тем связь теснее.

Отбор факторов, включаемых в модель множественной регрессии

Методы отбора факторов. При построении моделей множественной регрессии одним из важнейших этапов является отбор факторов, воздействующих на результирующий признак. Обычно он происходит в два этапа:

I. Содержательный анализ факторов. Выделяют те факторы, которые существенно влияют на результат.

II. Расчет количественных оценок, позволяющих оценить влияние факторов на результат и друг на друга. На их основании проводится окончательный отбор факторов. Здесь могут использоваться различные **методы отбора факторов**, например:

- а) использование парных коэффициентов корреляции;
- б) использование парных частных коэффициентов корреляции;
- в) расчет вкладов факторов в объясненную дисперсию;
- г) и т.п.

Рассмотрим более подробно использование парных коэффициентов корреляции.

Матрица парных коэффициентов линейной корреляции. Пусть в модели множественной регрессии предполагается использовать n признаков-факторов x_1, x_2, \dots, x_n , а y – результативный признак. Тогда для этих переменных может быть построена **матрица парных коэффициентов линейной корреляции**, или корреляционная матрица, которая по своей сути представляет совокупность коэффициентов корреляции между всеми возможными парами признаков:

$$\begin{array}{c} \begin{array}{ccccc} & y & x_1 & x_2 & \dots & x_n \\ y & \left(\begin{array}{ccccc} 1 & r_{x_1y} & r_{x_2y} & \dots & r_{x_ny} \\ r_{x_1y} & 1 & r_{x_1x_2} & \dots & r_{x_1x_n} \\ r_{x_2y} & r_{x_1x_2} & 1 & \dots & r_{x_2x_n} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_ny} & r_{x_1x_n} & r_{x_2x_n} & \dots & 1 \end{array} \right) \end{array} \end{array}$$

Эта матрица всегда симметрична относительно главной диагонали, а члены матрицы, стоящие на этой диагонали, равны 1, поскольку представляют собой корреляцию признака с самим собой.

Анализ корреляционной матрицы позволяет выявить факторы, тесно связанные между собой, т.е. коллинеарные.

Коллинеарность факторов – это тесная линейная связь между двумя факторами. Считается, что переменные явно коллинеарны, если коэффициент корреляции между ними по модулю превышает 0,7 .

При наличии коллинеарных признаков один из них нужно исключить из модели, чтобы между оставшимися факторами не было тесных связей. Разумеется, здесь речь идет именно о связи между признаками-факторами x_1, x_2, \dots, x_n . Коэффициенты корреляции между результатом y и факторами $x_1,$

x_2, \dots, x_n должны быть как можно ближе к единице по модулю, т.е. связь между ними должна быть тесной.

Мультиколлинеарность факторов – это тесная линейная связь между несколькими переменными, коллинеарность многих факторов.

По парным коэффициентам корреляции можно заметить лишь явную коллинеарность факторов. Чтобы оценить мультиколлинеарность всех факторов, имеет смысл построить матрицу парных коэффициентов корреляции между факторами (**матрицу межфакторной корреляции**) и рассчитать ее определитель. Эта матрица может быть легко получена из корреляционной матрицы вычеркиванием первых строки и столбца (соответствующих признаку-результату).

Очевидно, что новая матрица в случае полного отсутствия корреляции между факторами будет единичной матрицей (все недиагональные элементы равнялись бы нулю), и ее определитель равнялся бы единице. В противоположном случае, т.е. если бы между факторами была полная линейная зависимость, все элементы новой матрицы равнялись бы 1, и ее определитель был бы равен 0. Таким образом, чем ближе определитель такой матрицы к нулю, тем сильнее мультиколлинеарность факторов.

Мультиколлинеарность факторов отрицательно сказывается на качестве модели, поскольку:

- 1) из-за связи между факторами затрудняется оценка влияния отдельных факторов на результат, что затрудняет интерпретацию параметров регрессии,
- 2) может привести к включению в модель лишних параметров;
- 3) уменьшается точность оценок коэффициентов регрессии, растет дисперсия оценок и стандартные ошибки;
- 4) завышается коэффициент множественной корреляции (см. далее).

Требованиям к факторам, включаемым в модель множественной регрессии. Таким образом, основными требованиями к факторам, включаемым в модель множественной регрессии, являются следующие:

1) их существенное влияние на результативный признак;

2) отсутствие мультиколлинеарности;

3) количественная измеримость факторов. Если в модель включается качественный фактор, для него необходимо разработать количественную шкалу измерения, например, балльную, либо использовать фиктивные переменные (см. далее).

Коэффициент множественной корреляции. Тесноту линейной связи между результатом и всеми признаками-факторами можно измерить с помощью множественного коэффициента корреляции, который для линейной модели можно рассчитать по формуле:

$$R_{y x_1 x_2 \dots x_n} = \sqrt{1 - \frac{\Delta_{\text{м.п.к.к.}}}{\Delta_{\text{м.м-ф.к.}}}},$$

где $\Delta_{\text{м.п.к.к.}}$ – определитель матрицы парных коэффициентов корреляции;

$\Delta_{\text{м.м-ф.к.}}$ – определитель матрицы межфакторной корреляции.

Этот показатель будет более подробно рассмотрен в дальнейшем при изучении оценки качества модели.

Фиктивные переменные

Обычно в качестве признаков-факторов выступают переменные, принимающие количественные значения. Однако иногда бывает необходимо включить в модель качественные факторы, обычно измеряемые в номинальной шкале (шкале наименований). Это может быть пол, род занятий, образование, сезон и т.п. В шкале наименований нельзя производить

арифметические действия, и задано только отношение тождества (объект либо принадлежит некоторому множеству, либо нет).

Помимо номинальной шкалы, существуют также другие виды шкал, в которых можно осуществлять далеко не все вычисления (например, порядковая шкала, на которой заданы только отношения тождества и больше-меньше; шкала разностей, на которой можно вычитать и складывать, но умножать и делить нельзя и т.п.). Факторы, измеряемые в этих шкалах, тоже могут включаться в эконометрическую модель.

Чтобы ввести такие переменные в модель, необходимо поставить им в соответствие некоторые числа, с которыми удобно производить вычисления в абсолютной шкале. Построенные таким образом переменные называют **фиктивными переменными**.

Остальные переменные модели, в противоположность фиктивным, иногда называют **значащими**.

Фиктивные переменные помогают отразить в модели неоднородность структуры наблюдений по некоторому качественному признаку.

Чаще всего в качестве фиктивных переменных используются так называемые бинарные (булевы, дихотомические) переменные, которые могут принимать всего два значения - 0 или 1.

Например, введем фиктивную переменную d , обозначающую пол респондента:

$$d = \begin{cases} 1 - \text{мужской пол} \\ 0 - \text{женский пол} \end{cases}$$

Пусть эконометрическая модель отражает линейную зависимость спроса на некоторый продукт от цены на него в виде парной регрессии

$$y = ax + b + \varepsilon,$$

где y – спрос, как результирующий признак;

x – цена на продукт (признак-фактор);

ε – случайная компонента;

a и b – параметры модели.

Предположим, что зависимость спроса на этот продукт, кроме того, зависит еще и от пола предполагаемого покупателя, причем уравнения отличаются только свободным членом, т.е. $y = ax + b_1 + \varepsilon$ для мужчин и для $y = ax + b_2 + \varepsilon$ женщин. Эти два уравнения можно представить в виде одного уравнения множественной регрессии с двумя признаками-факторами – x и d :

$$y = ax + db_1 + (1-d)b_2 + \varepsilon = ax + d(b_1 - b_2) + b_2 + \varepsilon$$

В других случаях введение фиктивной переменной может отражать влияние неоднородности наблюдений не только на свободный член, но и на другие параметры регрессии.

Кроме того, может использоваться несколько фиктивных переменных. Например, можно ввести фиктивные переменные d_1 и d_2 :

$$d_1 = \begin{cases} 1 - \text{наличие высшего образования} \\ 0 - \text{отсутствие высшего образования} \end{cases}$$
$$d_2 = \begin{cases} 1 - \text{наличие стажа работы} \\ 0 - \text{отсутствие стажа работы} \end{cases}$$

Эти переменные можно перемножать между собой. Произведение d_1d_2 будет служить признаком одновременного наличия стажа и высшего образования (только в этом случае оно будет равно 1; если хотя бы одно условие отсутствует, $d_1d_2 = 0$).

Фиктивная переменная не обязательно должна быть бинарной. Если она используется для отражения в модели качественного признака, принимающего не два, а большее количество значений в номинальной шкале, можно каждому такому значению поставить в соответствие значение фиктивной переменной. Но на практике это делают редко, поскольку в этом случае сложно дать коэффициентам регрессии содержательную интерпретацию.

Например, если фиктивная переменная соответствует одному из четырех сезонов, она могла бы принимать значения 1, 2, 3 и 4, или любые другие четыре разных значения. В этом случае переменная не была бы бинарной. Однако обычно вводят не одну, а три переменных:

$$d_1 = \begin{cases} 1 - \text{зима} \\ 0 - \text{не зима} \end{cases} \quad d_2 = \begin{cases} 1 - \text{весна} \\ 0 - \text{не весна} \end{cases} \quad d_3 = \begin{cases} 1 - \text{лето} \\ 0 - \text{не лето} \end{cases}$$

Четвертая переменная не вводится, поскольку если бы была введена аналогичным образом переменная d_4 , то всегда выполнялась бы линейная зависимость между признаками факторами $d_1 + d_2 + d_3 + d_4 = 1$. Такая зависимость лишит исследователя возможности найти параметры регрессии с помощью метода наименьших квадратов, поскольку нарушится одна из его важных предпосылок (см. далее).

Пусть $y = a_1d_1 + a_2d_2 + a_3d_3 + b + \varepsilon$, где y – спрос на продукцию, зависящий от сезона. Тогда смысл параметров регрессии легко интерпретировать. В самом деле, тогда зимой значение спроса будет $a_1 + b$, весной $a_2 + b$, летом $a_3 + b$, осенью b . Каждый из коэффициентов a_1, a_2, a_3 представляет собой отклонение спроса в данном сезоне от осеннего спроса b .

В эконометрических моделях, отражающих зависимость результата от времени, т.е. регрессионных моделях с временными рядами, принято использовать три основных типа фиктивных переменных:

1) индикаторы принадлежности наблюдения к определенному периоду (для наблюдений от и до определенного момента времени они равны 1, а для всех остальных – нулю). Такие переменные используются для моделирования скачкообразных сдвигов в структуре наблюдений. Например, если предположить, что в деятельности экономической системы наблюдалась определенная тенденция в период с 2000 по 2007 гг., а до и после этих лет она резко отличалась, то имеет смысл использовать в модели переменную,

которая будет принимать единичные значения только для наблюдений из этого периода.

2) сезонные переменные — индикаторы принадлежности наблюдений к определенному сезону (месяцу, кварталу). Чаще всего используются при исследовании экономических явлений, имеющих четкие различия в своем сезонном протекании (например, моделирование спроса на зимнюю одежду и обувь).

3) линейный временной тренд. Здесь фиктивная переменная по сути своей представляет собой номер наблюдения. Она показывает, какой промежуток времени прошел от условного начала отсчета времени (нулевого момента) до того момента, к которому относится данное наблюдение.

Линейное уравнение множественной регрессии

В общем виде линейное уравнение множественной регрессии можно записать следующим образом:

$$y = a_1x_1 + a_2x_2 + \dots + a_nx_n + b + \varepsilon,$$

где y – результативный признак (зависимая, результирующая, эндогенная переменная);

n – число факторов, включенных в модель;

x_1, x_2, \dots, x_n – признаки-факторы (регрессоры, объясняющие, предикторные, предопределенные, экзогенные переменные);

a_1, a_2, \dots, a_n – коэффициенты регрессии;

b – свободный член регрессии;

ε – компонента, отражающая в модели влияние случайных факторов, из-за которых реальное значение показателя может отклоняться от теоретического (регрессионный остаток).

По своей природе результирующая переменная всегда случайна. Регрессионный остаток позволяет отразить в модели стохастическую, вероятностную природу экономических процессов. Кроме того, можно также сказать, что он отражает все прочие не учтенные в явном виде факторы, которые могут повлиять на результат.

В дальнейшем в этом разделе, рассматривая способы построения уравнения регрессии, случайную компоненту пока не будем учитывать, т.е. будем рассматривать только детерминированную часть результата.

Экономический смысл параметров регрессии. Коэффициенты и свободный член регрессии принято также называть параметрами регрессии, или параметрами модели.

Коэффициенты регрессии a_1, a_2, \dots, a_n , как видно из записи модели, представляют собой частные производные результата по отдельным признакам-факторам:

$$a_j = \frac{\partial y}{\partial x_j}; j = \overline{1, n}$$

Они показывают, на сколько изменяется результативный признак при изменении соответствующего признака на единицу и неизменных значениях остальных признаков. Поэтому иногда коэффициент линейной регрессии называют также предельной эффективностью фактора.

Знак коэффициента линейной регрессии всегда совпадает со знаком коэффициента корреляции, так как положительная корреляция означает, что результат растет с ростом фактора, а отрицательная – что с ростом фактора результат убывает.

Однако, сравнение коэффициентов регрессии при различных признаках-факторах между собой представляется затруднительным, поскольку различные факторы обычно имеют разные единицы измерения, характеризуются различными значениями средних и показателями вариации. Чтобы решить эту проблему, рассчитывают **стандартизованные коэффициенты регрессии** (см. далее). В отличие от стандартизованных коэффициентов регрессии коэффициенты регрессии a_1, a_2, \dots, a_n принято называть **коэффициентами чистой регрессии**.

Свободный член регрессии b показывает значение признака-результата при условии, что все признаки-факторы равны нулю. Если такая ситуация невозможна, свободный член может и не иметь экономического содержания.

Частные уравнения регрессии. На основе линейного уравнения множественной регрессии могут быть получены частные уравнения регрессии, в которых все факторы, кроме обычно одного, закреплены на своем среднем уровне. Такое частное уравнение регрессии устанавливает

связь между результативным признаком и одним из признаков-факторов при условии, что остальные факторы приравнены к своим средним значениям. Система таких уравнений выглядит следующим образом:

$$\begin{cases} y_{x_1 * x_2, x_3, \dots, x_n} = a_1 x_1 + a_2 \bar{x}_2 + \dots + a_n \bar{x}_n + b \\ y_{x_2 * x_1, x_3, \dots, x_n} = a_1 \bar{x}_1 + a_2 x_2 + \dots + a_n \bar{x}_n + b \\ \dots \\ y_{x_n * x_2, x_3, \dots, x_{n-1}} = a_1 \bar{x}_1 + a_2 \bar{x}_2 + \dots + a_n x_n + b \end{cases},$$

где $\bar{x}_j, j = \overline{1, n}$ - средние значения признаков-факторов.

Подстановкой в формулу (1.12) конкретных значений средних можно получить следующую систему линейных уравнений:

$$\begin{cases} y_{x_1 * x_2, x_3, \dots, x_n} = a_1 x_1 + B_1 \\ y_{x_2 * x_1, x_3, \dots, x_n} = a_2 x_2 + B_2 \\ \dots \\ y_{x_n * x_1, x_2, \dots, x_{n-1}} = a_n x_n + B_n \end{cases},$$

где

$$\begin{cases} B_1 = a_2 \bar{x}_2 + a_3 \bar{x}_3 + \dots + a_n \bar{x}_n + b \\ B_2 = a_1 \bar{x}_1 + a_3 \bar{x}_3 + \dots + a_n \bar{x}_n + b \\ \dots \\ B_n = a_1 \bar{x}_1 + a_2 \bar{x}_2 + \dots + a_{n-1} \bar{x}_{n-1} + b \end{cases}$$

Кроме того, можно построить частные уравнения регрессии и для нескольких независимых переменных, т.е. закрепить на среднем уровне все факторы, кроме нескольких.

На основе частных уравнений регрессии могут быть построены так называемые **частные коэффициенты эластичности** \mathcal{E}_i , которые

рассчитываются по формулам $\mathcal{E}_i = a_i \frac{\bar{x}_i}{y}$ и показывают, на сколько

процентов изменится результат при изменении фактора x_i на 1%. Расчет этих коэффициентов позволяет оценить, какие факторы более сильно воздействуют на результативный признак. Таким образом, их тоже можно использовать при отборе факторов в регрессионную модель.

Стандартизованное уравнение регрессии. Перейдем от переменных модели y, x_1, x_2, \dots, x_n к так называемым **стандартизованным переменным** по следующим формулам:

$$t_y = \frac{y - \bar{y}}{\sigma(y)}, t_{x_j} = \frac{x_j - \bar{x}_j}{\sigma(x_j)}, j = \overline{1, n},$$

где $\bar{y}, \bar{x}_j, j = \overline{1, n}$ - средние значения признаков;

$\sigma(y), \sigma(x_j), j = \overline{1, n}$ - средние квадратические отклонения признаков.

Для новых переменных среднее значение равно нулю, а среднее квадратическое отклонение равно единице.

Стандартизованное уравнение регрессии (или уравнение регрессии в стандартизованном масштабе) строится следующим образом:

$$t_y = \alpha_1 \cdot t_{x_1} + \alpha_2 \cdot t_{x_2} + \dots + \alpha_n \cdot t_{x_n},$$

где $t_y, t_{x_1}, \dots, t_{x_n}$ - стандартизованные переменные;

$\alpha_1, \alpha_2, \dots, \alpha_n$ – стандартизованные коэффициенты регрессии.

Для нахождения стандартизованных коэффициентов используют матрицу парных коэффициентов корреляции. Можно доказать, что для стандартизованных коэффициентов регрессии выполняется следующая система уравнений:

[illegible]

Она представляет собой систему n уравнений для n переменных, и ее можно использовать для определения стандартизованных коэффициентов регрессии. Эти коэффициенты можно сравнивать друг с другом и соответственно сравнивать различные факторы по силе воздействия на результат.

Можно доказать, что для линейной модели формулу для расчета коэффициента множественной корреляции можно преобразовать в формулу, основанную на использовании стандартизованных коэффициентов регрессии:

$$\mathbf{R}_{y x_1 x_2 \dots x_n} = \sqrt{\sum_{i=1}^n \alpha_i} \cdot \mathbf{r}_{y x_i},$$

где α_i – стандартизованные коэффициенты регрессии,

r_{yx_i} - парные коэффициенты корреляции результата с каждым из факторов.

Подставив в стандартизованное уравнение регрессии вместо стандартизованных переменных те формулы, по которым они рассчитывались, можно вернуться к уравнению чистой регрессии:

$$\frac{y - \bar{y}}{\sigma(y)} = \alpha_1 \cdot \frac{x_1 - \bar{x}_1}{\sigma(x_1)} + \alpha_2 \cdot \frac{x_2 - \bar{x}_2}{\sigma(x_2)} + \dots + \alpha_n \cdot \frac{x_n - \bar{x}_n}{\sigma(x_n)};$$

$$\frac{y}{\sigma(y)} - \frac{\bar{y}}{\sigma(y)} = \frac{\alpha_1}{\sigma(x_1)} \cdot x_1 - \alpha_1 \cdot \frac{\bar{x}_1}{\sigma(x_1)} + \frac{\alpha_2}{\sigma(x_2)} \cdot x_2 - \alpha_2 \cdot \frac{\bar{x}_2}{\sigma(x_2)} + \dots + \frac{\alpha_n}{\sigma(x_n)} \cdot x_n - \alpha_1 \cdot \frac{\bar{x}_n}{\sigma(x_n)};$$

$$\frac{y}{\sigma(y)} = \frac{\alpha_1}{\sigma(x_1)} \cdot x_1 + \frac{\alpha_2}{\sigma(x_2)} \cdot x_2 + \dots + \frac{\alpha_n}{\sigma(x_n)} \cdot x_n - \alpha_1 \cdot \frac{\bar{x}_1}{\sigma(x_1)} - \alpha_2 \cdot \frac{\bar{x}_2}{\sigma(x_2)} - \dots - \alpha_n \cdot \frac{\bar{x}_n}{\sigma(x_n)} + \frac{\bar{y}}{\sigma(y)}$$

Умножим обе части на $\sigma(y)$, получим:

$$y = \frac{\sigma(y) \cdot \alpha_1}{\sigma(x_1)} \cdot x_1 + \frac{\sigma(y) \cdot \alpha_2}{\sigma(x_2)} \cdot x_2 + \dots + \frac{\sigma(y) \cdot \alpha_n}{\sigma(x_n)} \cdot x_n - \sigma(y) \cdot \left(\alpha_1 \cdot \frac{\bar{x}_1}{\sigma(x_1)} + \alpha_2 \cdot \frac{\bar{x}_2}{\sigma(x_2)} - \dots + \alpha_n \cdot \frac{\bar{x}_n}{\sigma(x_n)} \right) + \bar{y}$$

Отсюда:

$$a_j = \frac{\sigma(y) \cdot \alpha_j}{\sigma(x_j)}$$

$$b = \bar{y} - \sigma(y) \cdot \left(\alpha_1 \cdot \frac{\bar{x}_1}{\sigma(x_1)} + \alpha_2 \cdot \frac{\bar{x}_2}{\sigma(x_2)} - \dots + \alpha_n \cdot \frac{\bar{x}_n}{\sigma(x_n)} \right)$$

Метод наименьших квадратов (МНК)

Оценка параметров линейных уравнений регрессии

Процедура построения системы нормальных уравнений и исходное соотношение, используемое в МНК. Для определения параметров функции, используемой в эконометрической модели, разработаны различные методы, наиболее простым и известным из которых является **метод наименьших квадратов (МНК)**. Рассмотрим суть этого метода на примере парной линейной регрессии.

Применение МНК к парной линейной регрессии. Итак, необходимо определить параметры a и b для функции $y = f(x) = ax + b$. Пусть значения показателей y и x измерялись n раз, т.е. имеются значения показателей y_1, y_2, \dots, y_n и x_1, x_2, \dots, x_n , всего n пар значений обоих показателей. Определим сумму квадратов отклонений фактических значений признака-результата y_i от значений, подсчитанных по уравнению регрессии $f(x_i)$:

$$\sum_{i=1}^n (f(x_i) - y_i)^2 = \sum_{i=1}^n (ax_i + b - y_i)^2$$

Чтобы построенная модель была как можно ближе к реальности, эта сумма должна быть как можно меньше. Отметим, что полученная сумма представляет собой функцию от двух переменных a и b , и чтобы найти ее минимум, приравняем к нулю ее частные производные по a и по b :

$$\sum_{i=1}^n (ax_i + b - y_i)^2 = \varphi(a, b) \rightarrow \min$$

$$\begin{cases} \frac{\partial \varphi}{\partial b} = 2 \sum_{i=1}^n (ax_i + b - y_i) = 0 \\ \frac{\partial \varphi}{\partial a} = 2 \sum_{i=1}^n x_i (ax_i + b - y_i) = 0 \end{cases}$$

$$\begin{cases} a \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \end{cases}$$

Итак, получено два линейных уравнения с двумя неизвестными – а и b (система линейна относительно параметров регрессии). Такую систему называют **системой нормальных уравнений**. Решив ее, можно определить искомые параметры.

Применение МНК к множественной линейной регрессии. Если необходимо определить параметры множественной линейной регрессии, т.е. параметры функции $y = f(x_1, x_2, \dots, x_m) = a_1x_1 + a_2x_2 + \dots + a_mx_m + b$, имея в запасе n наблюдений для каждого признака-фактора и для признака результата, то можно аналогичным образом получить систему нормальных уравнений, состоящую из (m + 1) линейного уравнения для (m + 1) неизвестной a_1, a_2, \dots, a_m, b :

$$\begin{cases} nb + a_1 \sum x_1 + a_2 \sum x_2 + \dots + a_m \sum x_m = \sum y \\ b \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 + \dots + a_m \sum x_1 x_m = \sum y x_1 \\ \dots \\ b \sum x_m + a_1 \sum x_1 x_m + a_2 \sum x_m x_2 + \dots + a_m \sum x_m^2 = \sum y x_m \end{cases},$$

где

$$\sum y = \sum_{i=1}^n y_i; \sum x_j = \sum_{i=1}^n x_{ji}, \sum x_j x_k = \sum_{i=1}^n x_{ji} x_{ki}, \sum y x_j = \sum_{i=1}^n y_i x_{ji}, k, j = \overline{1, m},$$

y_i, x_{ji} - i-ые значения наблюдаемых показателей (для каждого показателя их n).

Отметим, что систему уравнений для нахождения стандартизованных коэффициентов регрессии, которую мы рассматривали на лекции 31, также

получают путем применения МНК к стандартизированному уравнению регрессии и преобразования полученных выражений.

Методы решения системы нормальных уравнений. Решение построенной системы может быть осуществлено различными способами:

1) методом Гаусса, который заключается в том, что матрицу коэффициентов в уравнениях поэтапно преобразуют в единичную матрицу путем линейных преобразований этих уравнений (разрешающее уравнение делят на разрешающий элемент, получая на его месте единицу, а из всех остальных уравнений вычитают преобразованное разрешающее, умноженное на те коэффициенты, которые стоят в этих уравнениях в разрешающем столбце, с целью получить на их месте нули);

2) методом Крамера, который заключается в том, что рассчитывают определитель матрицы коэффициентов в уравнениях, а затем рассчитывают частные определители, поочередно заменяя один из столбцов в этой матрице столбцом свободных членов; значения переменных равны отношению соответствующих частных определителей к определителю первой матрицы;

3) методом обратной матрицы и т.д.

Матричная форма МНК

Рассмотрим систему нормальных уравнений МНК, используя обозначения матричной алгебры. А именно, введем следующие обозначения:

$$X = \begin{pmatrix} 1 & x_{1_1} & x_{2_1} & \dots & x_{m_1} \\ 1 & x_{1_2} & x_{2_2} & \dots & x_{m_2} \\ 1 & x_{1_3} & x_{2_3} & \dots & x_{m_3} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1_n} & x_{2_n} & \dots & x_{m_n} \end{pmatrix}; Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}; A = \begin{pmatrix} b \\ a_1 \\ a_2 \\ \dots \\ a_m \end{pmatrix},$$

где m – число признаков-факторов,

n - число наблюдений.

Каждая строка матрицы соответствует одному из наблюдений, а каждый столбец, кроме первого, - одному из факторов.

Если транспонировать матрицу X размерности $n \times (m + 1)$, в полученной матрице X^T размерности $(m + 1) \times n$ каждый столбец будет соответствовать одному из факторов, а строки - наблюдениям. Перемножив полученную матрицу X^T на X , получим симметричную матрицу размерности $(m + 1) \times (m + 1)$:

$$X^T X = \begin{pmatrix} n & \sum_{i=1}^n X_{1_i} & \sum_{i=1}^n X_{2_i} & \dots & \sum_{i=1}^n X_{m_i} \\ \sum_{i=1}^n X_{1_i} & \sum_{i=1}^n X_{1_i}^2 & \sum_{i=1}^n X_{1_i} X_{2_i} & \dots & \sum_{i=1}^n X_{1_i} X_{m_i} \\ \sum_{i=1}^n X_{2_i} & \sum_{i=1}^n X_{1_i} X_{2_i} & \sum_{i=1}^n X_{2_i}^2 & \dots & \sum_{i=1}^n X_{2_i} X_{m_i} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^n X_{m_i} & \sum_{i=1}^n X_{1_i} X_{m_i} & \sum_{i=1}^n X_{2_i} X_{m_i} & \dots & \sum_{i=1}^n X_{m_i}^2 \end{pmatrix}$$

Тогда система уравнений примет вид $X^T X A = X^T Y$. Умножим слева обе части этого выражения на матрицу $(X^T X)^{-1}$, получим: $(X^T X)^{-1} X^T X A = (X^T X)^{-1} X^T Y$. Поскольку $(X^T X)^{-1} X^T X = I$ (где I - единичная матрица), формула для нахождения вектора параметров A примет вид:

$$A = (X^T X)^{-1} X^T Y$$

МНК можно применять также, используя нелинейные функции. Он достаточно легко обобщается на полиномы более высоких степеней и ряд других функций.

Виды уравнений регрессии, параметры которых можно определить при помощи МНК:

- 1) линейная регрессия
- 2) нелинейная регрессия
 - а) полиномы различных степеней (квадратическая, кубическая функция и т.д.)
 - б) функции, преобразуемые к линейному виду (гиперболическая, степенная, экспоненциальная и некоторые другие функции).

Рассмотрим это более подробно.

Применение МНК к полиномиальной регрессии рассмотрим на примере использования квадратической функции в парной регрессии, т.е. при определении параметров уравнения $y = ax^2 + bx + c$. В этом случае будет построена система трех нормальных уравнений, с помощью которой можно найти три неизвестных параметра a , b и c (n – число наблюдений):

$$\begin{cases} a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i + nc = \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_i^3 + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ a \sum_{i=1}^n x_i^4 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i^2 y_i \end{cases}$$

Однако, такой способ расчетов достаточно сложен из-за высоких степеней, в которые придется возводить значения фактора. Поэтому лучше применять методику, которая будет рассмотрена далее.

Многие нелинейные функции можно подвергнуть **линеаризации**, т.е. преобразовать в линейные различными способами, после чего к ним также применяется МНК. Рассмотрим несколько примеров.

Применение МНК к гиперболической функции. Эту функцию можно линеаризовать путем замены переменных. В уравнении гиперболы $y = a/x + b$ осуществим замену переменных следующим образом: $z = 1/x$. После этого уравнение станет линейным: $y = az + b$.

Применение МНК к степенной функции. Линеаризацию степенной функции $y = ax^b$ осуществляют, логарифмируя обе части уравнения: $\ln y = \ln a + b \cdot \ln x$. Можно осуществить замену переменных $z = \ln x$ и $q = \ln y$, тогда уравнение примет вид $q = \ln a + bz$ с неизвестными параметрами $\ln a$ и b , которые можно найти с помощью МНК.

Применение МНК к экспоненциальной функции ($y = ae^{bx}$). Эта функция также линеаризуется путем логарифмирования: $\ln y = \ln a + bx$; после замены переменной $q = \ln y$ получим $q = \ln a + bx$.

Вышеперечисленные приемы линеаризации легко обобщаются на случай множественной регрессии. Рассмотрим это на примере степенной функции $y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_m^{b_m}$. Логарифмируя это уравнение, получим: $\ln y = \ln a + b_1 \cdot \ln x_1 + b_2 \cdot \ln x_2 + \dots + b_m \cdot \ln x_m$, после чего путем замены переменных можно перейти к множественной линейной регрессии и применять МНК.

Линеаризация некоторых других функций для применения МНК. Например, рассмотрим функцию $y = f(x_1, x_2) = a_1 x_1 + a_2 x_1^2 + a_3 x_2^2 + b$. Осуществим замену переменных следующим образом: $x_3 = x_1^2$; $x_4 = x_2^2$. Тогда функция примет линейный вид: $y = a_1 x_1 + a_2 x_3 + a_3 x_4 + b$.

Свойства оценок, получаемых при помощи МНК

Понятно, что при многократном проведении наблюдений в результате расчетов будут получены различные значения параметров (в результате случайных колебаний). Например, если мы определяем параметры парной линейной регрессии a и b , то в результате исследования одной выборки мы получим значения параметров a_1 и b_1 , по другой выборке - a_2 и b_2 , и т.д. можем получить бесконечно много оценок параметров. Таким образом, сами оценки представляют собой случайную величину, для которой можно рассчитать вероятностные характеристики.

В теории вероятностей среднее значение случайной величины, полученное при неограниченно большом числе опытов, называют ее **математическим ожиданием** и обозначают $M(x)$ (x – случайная величина).

Математическое ожидание квадрата отклонения случайной величины от своего математического ожидания называют **дисперсией** (соответствует понятию дисперсии в статистике – средний квадрат отклонений от среднего) и обозначают $D(x)$ или $\sigma^2(x)$. Для расчета дисперсии удобно использовать не ее определение $D(x) = M(x - M(x))^2$, а следующую формулу: $D(x) = M(x^2) - M^2(x)$. Иными словами, дисперсию можно рассчитать, как разность между математическим ожиданием квадрата случайной величины и квадратом ее математического ожидания¹.

Итак, выборочные оценки параметров имеют математическое ожидание и дисперсию.

Если бы мы могли охватить в исследовании не выборку, а всю генеральную совокупность данных, то получили бы значения параметров регрессии, которые условно можно назвать истинными.

¹ Эта формула доказывается так же, как и аналогичная формула для выборочной дисперсии (дисперсия равна разности между средним квадратом значений показателя и квадратом среднего значения: $\sigma^2(x) = \overline{x^2} - \bar{x}^2$).

Оценки параметров, полученные МНК, обладают важными свойствами, строгое доказательство которых приводится в математической статистике (здесь не рассматривается):

- 1) несмещенность;
- 2) состоятельность;
- 3) эффективность.

Рассмотрим их подробно.

Несмещенность. Свойство несмещенности заключается в том, что математическое ожидание оценки равно неизвестному истинному значению параметра. Это означает, что выборочные оценки как бы концентрируются вокруг неизвестных истинных значений параметров. Это очень важное свойство, - если бы оно не выполнялось, метод давал бы заведомо неверную информацию.

Состоятельность. Свойство состоятельности заключается в том, что при стремлении числа наблюдений к бесконечности дисперсии оценок стремятся к нулю. Это означает, что с ростом числа наблюдений их разброс становится все меньше, оценки становятся все более надежными, все плотнее концентрируются вокруг истинных значений.

Эффективность. Свойство эффективности заключается в том, что эти оценки имеют наименьшую дисперсию по сравнению с любыми другими оценками параметров. Собственно, именно на этом и основан МНК (см. исходное соотношение (2.2)).

Практическая значимость перечисленных свойств заключается еще и в том, что с ростом объема выборки не происходит накопление регрессионных остатков.

Предпосылки МНК

Следует отметить, что вышеперечисленные свойства оценок МНК имеют место лишь при некоторых предположениях о регрессорах и случайной компоненте (регрессионном остатке) тренда. Перечислим их.

Перечень предпосылок МНК (условия Гаусса-Маркова):

1) математическое ожидание регрессионного остатка должно быть равно нулю ($M(\varepsilon) = 0$);

2) дисперсия регрессионного остатка должна быть постоянна (это свойство называется **гомоскедастичностью** остатка, слово складывается из двух частей: «гомо» - однородность и «скедастичность» - разброс, вариабельность) и конечна ($D(\varepsilon) = \text{const} < \infty$);

3) значения регрессионного остатка не должны зависеть друг от друга (т.е. не должно быть **автокоррелированности** остатков) ($\text{Cov}(\varepsilon_i, \varepsilon_j) \approx 0$, где – выборки значений случайной компоненты ε в любых двух наборах наблюдений);

4) регрессионный остаток и признаки не должны зависеть друг от друга ($\text{Cov}(\varepsilon, y) \approx 0$, $\text{Cov}(\varepsilon, x_j) \approx 0$, $\forall j$);

5) не должно быть мультиколлинеарности ($\text{Cov}(x_i, x_j) \approx 0$, $\forall i, j$).

Регрессионные остатки, для которых выполняются вышеперечисленные требования, представляют собой так называемый «белый шум», т.е. независимые друг от друга значения нормально распределенной случайной величины (более подробно рассматривается при изучении стационарных временных рядов).

Последствия нарушения предпосылок МНК. Рассмотрим, что может произойти при нарушении одной или нескольких из названных предпосылок.

1) Если в регрессионном уравнении присутствует свободный член, ожидаемое значение случайной компоненты всегда равно нулю (если бы это было не так, было бы достаточно просто пересчитать свободный член).

Нарушаться это требование может лишь в том случае, если по каким-либо причинам требуется, чтобы свободный член равнялся нулю или другому фиксированному значению. Тогда полученная с помощью модели оценка может оказаться смещенной.

2) **Гетероскедастичность**, т.е. отсутствие гомоскедастичности, может привести к тому, что оценки МНК не будут обладать свойством эффективности (само слово «гетероскедастичность» складывается из двух частей: «гетеро» - разнородность и «скедастичность» - вариабельность). Кроме того, хотя сами оценки параметров и останутся несмещенными, но стандартные ошибки этих оценок (они рассчитываются на основе дисперсии) могут оказаться смещены, что иногда приводит к неправильным результатам при проверке модели на значимость.

3) Наличие автокорреляции, как и гетероскедастичность, делает оценки неэффективными. Кроме того, оно тоже может привести к неправильному расчету стандартных ошибок модели и, как следствие, ненадежности проверки модели на значимость.

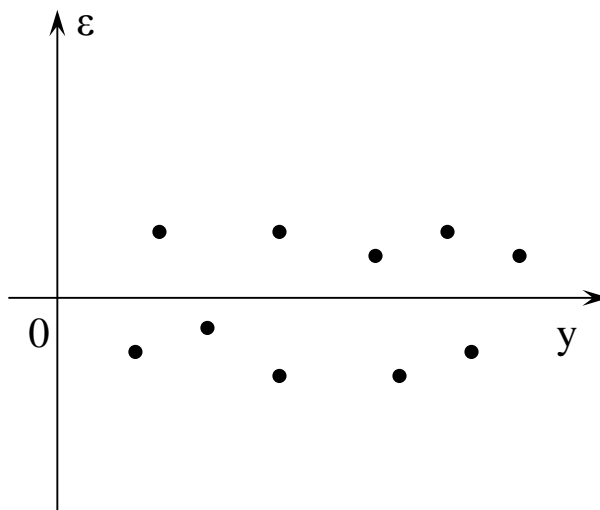
4) Если можно выявить зависимость между значениями регрессионного остатка и каким-либо из признаков, это говорит о том, что случайная компонента не является случайной по своей сути. В построенную модель необходимо внести исправления, учитывающие эту закономерность.

5) Отрицательные последствия мультиколлинеарности факторов были подробно рассмотрены ранее, а именно она затрудняет интерпретацию параметров регрессии, уменьшает точность оценок коэффициентов; приводит к росту стандартных ошибок и завышает коэффициент множественной корреляции.

Рассмотрим некоторые методы проверки выполнения предпосылок Гаусса-Маркова и приемы исследования в случаях, когда они нарушаются.

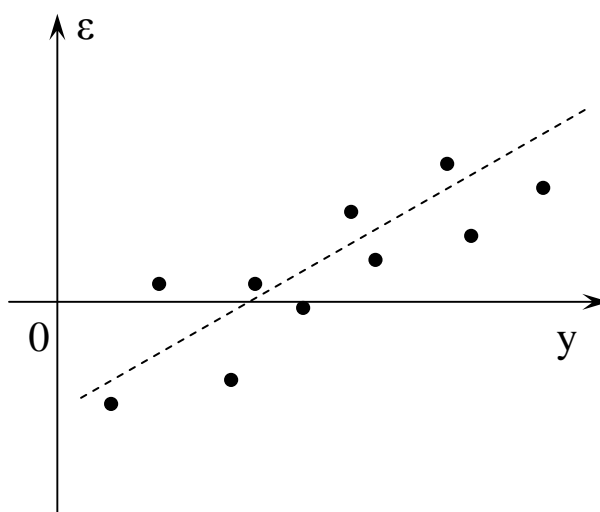
Способ проверки остатков на случайный характер

Для проверки остатков на случайный характер строят график зависимости случайной компоненты от значений результативного признака:



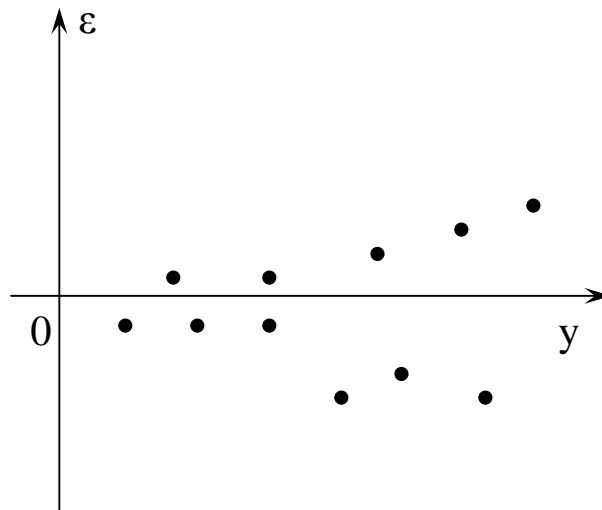
Если значения остатков расположены вблизи горизонтальной прямой (оси абсцисс), то их можно считать случайными (как на рисунке выше).

На рисунке ниже остатки носят систематический характер:



Зависимость регрессионного остатка от значений результативного признака (систематический характер остатков)

На следующем рисунке дисперсия остатков, соответствующих большим значениям y , больше, чем дисперсия при малых y , т.е. имеет место гетероскедастичность остатков:



Гетероскедастичность остатков

Кроме того, существует ряд специальных тестов, разработанных для проверки остатков на гомоскедастичность и отсутствие автокорреляции.

Наиболее известным тестом для проверки на гомоскедастичность является тест Голдфелда-Квандта, а для проверки остатков на автокорреляцию - тест Дарбина-Уотсона. Студентам рекомендуется изучить их самостоятельно.

Обобщенный МНК

Рассмотрим отдельно две важные предпосылки МНК: гомоскедастичность и отсутствие автокорреляции остатков. Взяв n наблюдений, для каждого из них можно получить регрессионный остаток $\epsilon_1, \epsilon_2, \dots, \epsilon_n$. Каждый из этих остатков сам по себе является случайной величиной. Для этих случайных величин можно построить ковариационную

матрицу, на диагонали которой будут стоять дисперсии остатков, а остальные элементы будут представлять собой ковариации между ними (матрица симметрична относительно главной диагонали):

$$\Omega = \begin{matrix} & \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \\ \begin{matrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{matrix} & \begin{pmatrix} \sigma^2(\varepsilon_1) & \text{Cov}(\varepsilon_1, \varepsilon_2) & \dots & \text{Cov}(\varepsilon_1, \varepsilon_n) \\ \text{Cov}(\varepsilon_1, \varepsilon_2) & \sigma^2(\varepsilon_2) & \dots & \text{Cov}(\varepsilon_2, \varepsilon_n) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(\varepsilon_1, \varepsilon_n) & \text{Cov}(\varepsilon_2, \varepsilon_n) & \dots & \sigma^2(\varepsilon_n) \end{pmatrix} \end{matrix}$$

Если остатки гомоскедастичны, то элементы на главной диагонали этой матрицы будут равны между собой. Если автокорреляция остатков отсутствует, то ненулевые элементы этой матрицы могут стоять только на главной диагонали. Существенное отличие любого другого элемента матрицы от нуля означает, что регрессионные остатки коррелируют.

Как уже было сказано, гетероскедастичность и автокорреляция остатков приводят к тому, что оценки, полученные МНК, будут неэффективными. Исключить то и другое можно с помощью модификации МНК – **обобщенного метода наименьших квадратов (ОМНК)**, суть которого сводится к тому, что при нахождении вектора параметров A используют не формулу, которую мы ранее получили для матричной формы МНК, а следующую формулу:

$$A = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y,$$

где Ω^{-1} – матрица, обратная ковариационной матрице Ω .

Можно доказать, что при использовании этой формулы оценки будут обладать свойством эффективности (теорема Айткена). Доказательство можно найти, например, в [Яновский Л.П., Буховец А.Г. Введение в эконометрику: уч. пособие – 2-е изд., доп. – М.: Кнорус, 2007. – 256 с.].

Исключение гетероскедастичности с помощью ОМНК

Предположим, что выполняется требование равенства математического ожидания регрессионного остатка нулю. Тогда дисперсия регрессионных остатков равна просто ожидаемому квадрату остатка: $\sigma^2(\varepsilon) = M(\varepsilon - M(\varepsilon))^2 = M(\varepsilon^2)$; $\sigma^2(\varepsilon_i) = M(\varepsilon_i^2)$.

Предположим, что требование отсутствия автокорреляции остатков тоже выполняется. Тогда ковариационная матрица остатков примет вид диагональной матрицы (ненулевые элементы стоят только на главной диагонали):

$$\Omega = \begin{matrix} & \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \\ \begin{matrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{matrix} & \left(\begin{array}{cccc} \sigma^2(\varepsilon_1) & 0 & \dots & 0 \\ 0 & \sigma^2(\varepsilon_2) & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2(\varepsilon_n) \end{array} \right) \end{matrix}$$

Пусть остатки гетероскедастичны, т.е. элементы на главной диагонали матрицы не равны между собой. Применение ОМНК с такой ковариационной матрицей сведется к тому, что в каждом i -м наблюдении все значения переменных будут поделены на одно и то же число $\sigma^2(\varepsilon_i)$. Такая модификация ОМНК называется **взвешенным МНК**.

Однако в реальных экономических задачах дисперсии регрессионных остатков для отдельных наблюдений неизвестны, и нет возможности построить ковариационные матрицы ни в каком виде. Поэтому вместо этих матриц обычно используют какую-либо их оценку.

Для определения коэффициентов при использовании взвешенного МНК может быть использован следующий подход. Предположим, что дисперсии остатков $\sigma^2(\varepsilon_i)$ пропорциональны величине $\sigma^2(\varepsilon)$ (дисперсии генеральной совокупности значений случайной компоненты). Коэффициенты пропорциональности обозначим K_i , - эти коэффициенты характеризуют

неоднородность дисперсии (способ их нахождения обсудим позже). Получим для каждого из n наблюдений:

$$\sigma^2(\varepsilon_i) = \sigma^2(\varepsilon) * K_i$$

В основе применения МНК к линеаризованной функции лежит соотношение (на примере парной линейной регрессии), которое может быть представлено следующим образом: $\sum_{i=1}^n (ax_i + b - y_i)^2 = \sum_{i=1}^n (\varepsilon_i)^2$. Если в левой части этого выражения каждое слагаемое в скобках разделить на $\sqrt{K_i}$, то в результате каждое слагаемое в правой части будет скорректировано на величину K_i . Поскольку из $\sigma^2(\varepsilon_i)/K_i = \sigma^2(\varepsilon)$, можно условно считать, что после такого преобразования данные будут гомоскедастичны, т.е. иметь общую дисперсию $\sigma^2(\varepsilon)$.

Итак, чтобы применить к парной линейной регрессии ОМНК в случае гетероскедастичности остатков, необходимо обе части уравнения $y = ax + b$ разделить на $\sqrt{K_i}$ для всех наблюдений:

$$y_i / \sqrt{K_i} = a x_i / \sqrt{K_i} + b / \sqrt{K_i}, \quad i = \overline{1, n}$$

Чтобы это сделать, исходные данные модели – значения x_i и y_i , делят на $\sqrt{K_i}$. Одновременно осуществляют замену переменных

$$\gamma_i = y_i / \sqrt{K_i}; \quad \alpha_i = x_i / \sqrt{K_i}; \quad \beta_i = 1 / \sqrt{K_i}, \quad i = \overline{1, n}$$

Значения новых переменных γ и α представляют собой значения показателей, взвешенные на коэффициенты $\beta_i = 1 / \sqrt{K_i}$. В общем случае эти веса надо задать для каждого наблюдения (каждой пары γ_i и α_i).

После такой замены уравнение регрессии примет вид

$$\gamma = a * \alpha + b * \beta$$

Полученное выражение представляет собой уравнение множественной (двухфакторной) линейной регрессии, в которой результирующий признак обозначен γ , а признаки-факторы - α и β . Параметры регрессии a и b можно найти из системы нормальных уравнений. В данном случае первое уравнение в системе следует опустить, так как свободный член регрессии здесь равен нулю (здесь оба параметра - a и b - представляют собой коэффициенты при переменных). Система примет вид:

$$\begin{cases} a \sum \alpha^2 + b \sum \alpha\beta = \sum \gamma\alpha \\ a \sum \alpha\beta + b \sum \beta^2 = \sum \gamma\beta \end{cases}$$

где

$$\sum \alpha^2 = \sum_{i=1}^n \alpha_i^2; \sum \beta^2 = \sum_{i=1}^n \beta_i^2; \sum \alpha\beta = \sum_{i=1}^n \alpha_i\beta_i; \sum \gamma\alpha = \sum_{i=1}^n \gamma_i\alpha_i; \sum \gamma\beta = \sum_{i=1}^n \gamma_i\beta_i$$

Каким образом определяются коэффициенты K_i ? Существуют различные подходы к их определению, и выбор любого из них неизбежно влияет на значение полученных параметров модели.

Иногда предполагают, что этими коэффициентами являются сами значения фактора. В многофакторной модели при этом одновременно встает проблема выбора одного из факторов (того, значения которого будут использованы при расчете весов). Например, можно взять последний по порядку фактор в множественной регрессии.

Следует отметить, что при этом, чем меньше значение фактора, тем на меньшую величину будет поделена величина дисперсии, т.е. весовой коэффициент $\frac{1}{\sqrt{K_i}}$ будет больше. Тем самым повышаются веса дисперсий ошибок в наблюдениях с меньшими значениями. Это говорит о том, что предположение о пропорциональности между коэффициентами K_i и значениями фактора может быть вполне обосновано с экономической точки

зрения: большим значениям фактора действительно может соответствовать большая дисперсия, которую необходимо умножить на меньший вес, чтобы добиться гомоскедастичности.

Исключение автокорреляции в остатках с помощью ОМНК

Рассмотрим случай автокорреляции остатков для модели, в которой наблюдения упорядочены во времени. Будем считать, что $M(\varepsilon) = 0$, и остатки гомоскедастичны.

Возьмем так называемый **авторегрессионный процесс первого порядка**, когда каждое последующее значение случайной компоненты связано с предыдущим линейной зависимостью:

$$\varepsilon_t = p\varepsilon_{t-1} + v_t,$$

где $t = 1, 2, \dots, n$ – номера последовательных наблюдений;

v_t - случайная компонента построенной зависимости¹, имеющая нулевое математическое ожидание и дисперсию σ_0^2 , не подверженная автокорреляции;

p - коэффициент авторегрессии.

Так как величины независимы, дисперсию суммы можно посчитать по следующей формуле (постоянный сомножитель выносим за скобки, возводя в квадрат, по свойству дисперсии):

$$D(\varepsilon_t) = p^2 D(\varepsilon_{t-1}) + D(v_t)$$

Поскольку остатки гомоскедастичны, $D(\varepsilon_t) = D(\varepsilon_{t-1}) = \sigma^2$, получим:

$$\sigma^2 = p^2 \sigma^2 + \sigma_0^2$$

$$\sigma^2 = \frac{\sigma_0^2}{1 - p^2}$$

¹ Обычно предполагается, что эта случайная величина имеет нормальное распределение.

Отсюда следует, что $|p| < 1$ (так как величина дисперсии должна быть положительной).

Найдем ковариацию² двух соседних остатков, подставляя вместо ε_t выражение $(p\varepsilon_{t-1} + v_t)$. При этом учтем, что математическое ожидание каждого из них равно нулю, и что математическое ожидание произведения независимых случайных величин ε_{t-1} и v_t , можно рассчитать, как произведение математических ожиданий:

$$\begin{aligned} \text{Cov}(\varepsilon_t, \varepsilon_{t-1}) &= M(\varepsilon_t * \varepsilon_{t-1}) - 0 = M(\varepsilon_{t-1} * (p\varepsilon_{t-1} + v_t)) = \\ &= M(p\varepsilon_{t-1}^2) + M(\varepsilon_{t-1}v_t) = p * D(\varepsilon_{t-1}) + M(\varepsilon_{t-1} * v_t) = p\sigma^2 \end{aligned}$$

Можно показать, что ковариации любой пары остатков рассчитываются по формуле:

$$\text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = p^k \sigma^2$$

Тогда ковариационная матрица примет вид:

$$\Omega = \sigma^2 \begin{matrix} & \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \\ \varepsilon_1 \left(\begin{array}{cccc} 1 & p & \dots & p^{n-1} \\ \varepsilon_2 & p & 1 & \dots & p^{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ \varepsilon_n & p^{n-1} & p^{n-2} & \dots & 1 \end{array} \right) \end{matrix}$$

Если параметр p известен, то для нахождения параметров линейной функции регрессии можно применить ОМНК.

Покажем, что при этом будет устранена автокорреляция остатков. Рассмотрим множественную линейную регрессию $y = a_1x_1 + a_2x_2 +$

² Для расчета ковариации в теории вероятностей можно использовать следующую формулу: $\text{Cov}(x, y) = M(x*y) - M(x)*M(y)$ (рекомендуется сравнить с формулой, используемой в статистике: $\text{Cov}(x, y) = \overline{xy} - \bar{x} * \bar{y}$).

+ ... + $a_m x_m$ + b + ε . Запишем уравнения регрессии для периодов t и $(t - 1)$, умножив обе части последнего уравнения на p :

$$y_t = a_1 x_{1t} + a_2 x_{2t} + \dots + a_m x_{mt} + b + \varepsilon_t$$

$$p y_{t-1} = p a_1 x_{1\ t-1} + p a_2 x_{2\ t-1} + \dots + p a_m x_{m\ t-1} + p b + p \varepsilon_{t-1}$$

Вычтем из первого уравнения второе, преобразовав результат к следующему виду:

$$y_t - p y_{t-1} = a_1 (x_{1t} - p x_{1\ t-1}) + a_2 (x_{2t} - p x_{2\ t-1}) + \dots + \\ + a_m (x_{mt} - p x_{m\ t-1}) + b(1 - p) + \varepsilon_t - p \varepsilon_{t-1}$$

Применив формулу ($\varepsilon_t = p \varepsilon_{t-1} + v_t$), получим

$$y_t - p y_{t-1} = a_1 (x_{1t} - p x_{1\ t-1}) + a_2 (x_{2t} - p x_{2\ t-1}) + \dots + \\ + a_m (x_{mt} - p x_{m\ t-1}) + b(1 - p) + v_t$$

В новой модели устранена автокорреляция остатков, так как новые остатки - v_t - независимы.

Для определения неизвестного параметра авторегрессии p можно использовать различные методы оценки. Проще всего оценить его с помощью обычного МНК, применяя его к уравнению авторегрессии остатков $\varepsilon_t = p \varepsilon_{t-1} + v_t$. Способ получения оценки дисперсии регрессионных остатков σ^2 будет рассмотрен позже. С помощью оценок p и σ^2 можно получить оценку ковариационной матрицы для применения ОМНК. Такой способ нахождения этой матрицы получил название **доступного ОМНК**.