Минестерство образования Республики Беларусь

УО «Витебский государственный технологический университет»

Тема7. «Математическая обработка результатов наблюдений»

Кафедра теоретической и прикладной математики.

разработана доц.Дуниной Е.Б.

7.1 Измерения и их погрешности.

Под измерениями понимают сравнение некоторой величины с другой (однородной) величиной, принятой за единицу.

Различают прямые и косвенные измерения.

При прямых измерениях исследуемая величина сравнивается с единицей измерения непосредственно или с помощью измерительного прибора (например, измерение длин линейкой)

При косвенных измерениях величины ее значения определяются по результатам прямых измерений других величин, связанных с рассматриваемой величиной заданной функциональной зависимостью (например, измерение плотности тела по измерениям его массы и объема).

В результате измерений получают приближенные значения величины, а не ее точное значение.

При измерениях погрешности неизбежны.

Погрешностью, или ошибкой, измерения называется разность $x - a$

между результатом измерения ***x*** **и точным значением** a **измеряемой величины.**

Погрешности измерений обычно неизвестны, т.к. неизвестно и точное значение измеряемой величины.

Различают следующие виды погрешностей измерений:

грубые ошибки – ошибки, сделанные вследствие неверной записи показаний прибора, неправильно прочитанного отсчета и т.д.;

систематические погрешности – погрешности, связанные с ограниченной точностью прибора, неправильной его установкой и т.д.

Систематические ошибки можно устранить путем введения соответствующих поправок.

Случайные погрешности вызываются большим числом случайных причин, действие которых на каждое измерение различно и заранее не может быть учтено.

Случайные ошибки являются неустранимыми.

В качестве закона распределения случайных погрешностей измерения чаще всего принимается нормальный закон распределения.

7.2 Оценка точного значения измеряемой величины.

Пусть в итоге *n* независимых измерений некоторой величины получены следующие результаты $x_1, x_2, x_3, \ldots, x_n$.

Будем предполагать, что эти результаты свободны от грубых и систематических ошибок.

Оценить точное значение a - это значит:

а) определить функцию $\alpha = \alpha(x_1, x_2, x_3, ..., x_n)$

которая обеспечивает достаточно близкое приближение к значению $a;$

б) указать границы интервала $(\alpha - \delta_1, a + \delta_2)$

который с заданной вероятностью γ

покрывает истинное значение $a.$

Оценка $\alpha = \alpha(x_1, x_2, x_3, ..., x_n)$

должна быть несмещенной, состоятельной и эффективной.

Как обычно вводят среднее арифметическое значение (среднее значение) $\overline{x}$ **результатов,**

среднее квадратическое отклонение этих результатов от их среднего значения и «исправленное» среднее квадратическое отклонение или эмпирический стандарт *s*

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i}{n},$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2},$$

$$s = \sqrt{\frac{n}{n-1}}\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Если все измерения проведены с одинаковой точностью, то в качестве оценки точного значения a **измеряемой величины**

принимают среднее арифметическое значение результатов

$$a = \bar{x} = \frac{\sum_{i=1}^{k} x_i}{n}.$$

Эта оценка является несмещенной и состоятельной.

Эта оценка относится к числу точечных.

В качестве интервальных оценок используют оценки вида

$$|\bar{x} - a| < \delta \quad (\delta > 0)$$ **или** $$\bar{x} - \delta < a < \bar{x} + \delta.$$

Величина δ **(точность оценки) определяется по заданной доверительной вероятности** γ **(надежности оценки);**

γ **обычно задается в виде одного из трех значений:**

$$\gamma = 0{,}95, \quad \gamma = 0{,}99, \quad \gamma = 0{,}999.$$

а) Доверительная оценка при известной точности измерений.

Если известно среднее квадратическое отклонение $\sigma,$

то доверительная оценка имеет вид $$|\bar{x} - a| < \frac{t\sigma}{\sqrt{n}}$$

где *n* – число измерений, а значение *t* определяется по заданной доверительной вероятности γ **из условия** $$2\Phi(t) = \gamma$$

и находится с помощью таблиц.

Точность оценки в этом случае выражается формулой

$$\delta = \frac{\sigma t}{\sqrt{n}}.$$

б) Доверительная оценка при неизвестной точности измерений.

Если **среднее квадратическое отклонение заранее неизвестно, то доверительная оценка принимает вид**

$$|\bar{x} - a| < \frac{t_\gamma s}{\sqrt{n}}.$$

Множитель t_γ **зависит от надежности** γ **и от числа измерений.**

Значение этого множителя находят по таблицам.

в) Правило трех сигм.

Поскольку надежность доверительной оценки выбирается заранее, на практике широко применяется правило трех сигм:

отклонение истинного значения от среднего арифметического результатов измерений не превосходит утроенной средней квадратической погрешности этого среднего значения

$$\left|a-\bar{x}\right| < \frac{3\sigma}{\sqrt{n}} \quad (7.1)$$

при известной величине $\sigma,$

$$\left|a-\bar{x}\right| < \frac{3s}{\sqrt{n}} \quad (7.2)$$

при неизвестной величине $\sigma,$

Оценка (7.1) имеет надежность

$$2\Phi(3) = 0{,}9973$$

(см.вопрос 4.4).

7.3 Метод наименьших квадратов.

При измерении двух величин *x* и *y* получены следующие данные

$\boldsymbol{x}$	x_1	x_2	$\cdots$	x_n
$\boldsymbol{y}$	y_1	y_2	$\cdots$	y_n

Известен также вид функциональной зависимости $$y = f(x).$$

Данная функция зависит от некоторых параметров значения которых требуется определить.

Значения *y*, полученные из формулы при заданных значениях x_i, **как правило, не совпадают с экспериментальными значениями, приведенными в таблице.**

Разности

$$f(x_i) - y_i = \varepsilon_i, \quad i = 1, \dots, n$$

называются уклонениями или погрешностями.

Требуется подобрать параметры функции *f(x)* так, чтобы уклонения ε_i **оказались наименьшими.**

Критерий, лежащий в основе метода наименьших квадратов:

параметры функции *f(x)* выбирают так, чтобы оказалась минимальной сумма квадратов уклонений:

$$\boxed{\sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 \geq 0.}$$ **!**

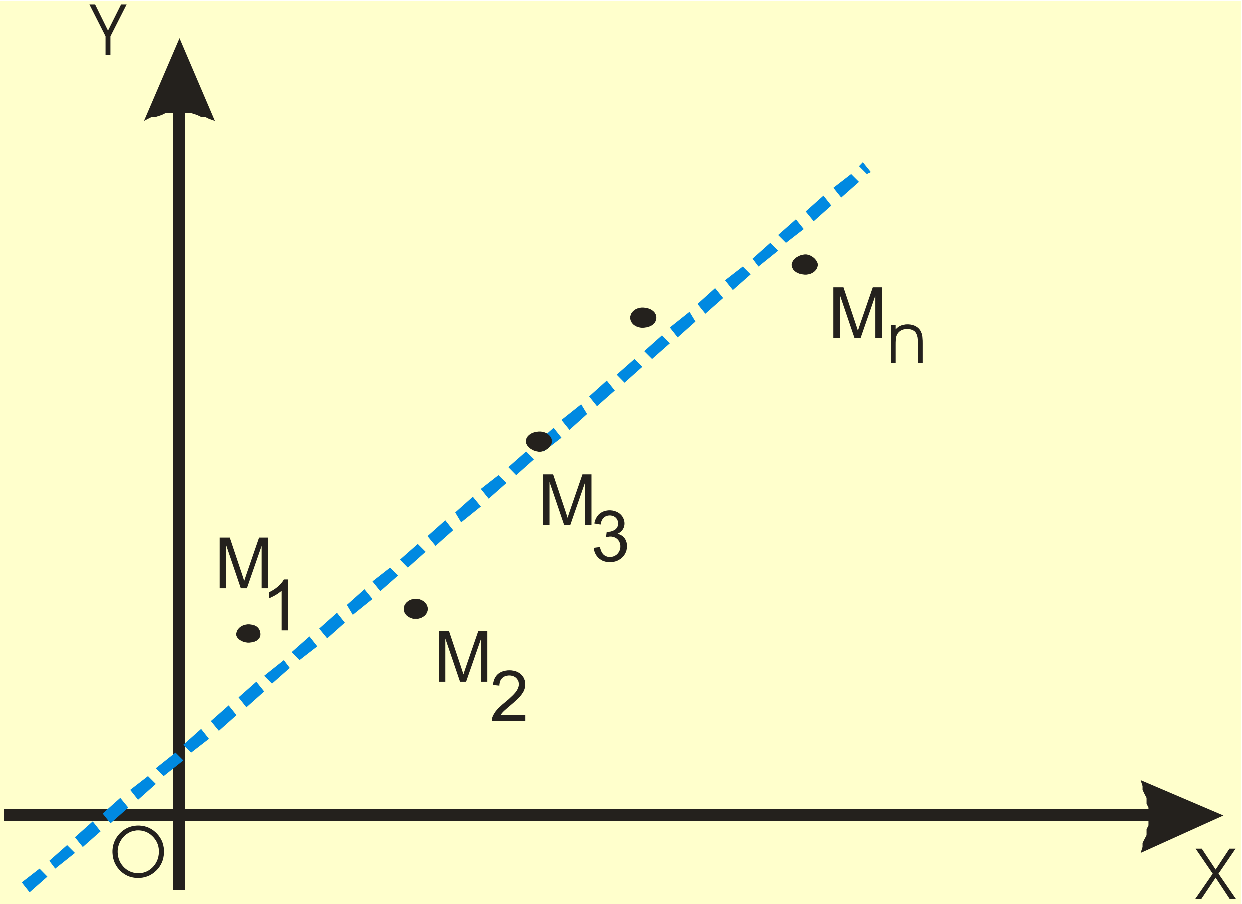
7.4 Определение параметров эмпирических формул в случае линейной зависимости.

Пусть необходимо установить зависимость между двумя величинами *x* и *y*, результаты измерений которых занесены в таблицу

$\boldsymbol{x}$	x_1	x_2	$\cdots$	x_n
$\boldsymbol{y}$	y_1	y_2	$\cdots$	y_n

Рассмотрим в декартовой системе координат *xOy* точки

$$M_1(x_1, y_1); M_2(x_2, y_2), \ldots, M_n(x_n, y_n)$$



Если эти точки почти лежат на некоторой прямой, естественно предположить, что между *x* и *y* существует линейная зависимость, выражающаяся формулой

$$y = ax + b, \qquad (7.3)$$

где a, b **- некоторые постоянные коэффициенты (параметры), подлежащие определению.**

Запишем выражение (7.3) в виде

$$ax + b - y = 0. \qquad (7.4)$$

Подставляя координаты точек

x_i и y_i $i = 1,\dots,n$

в формулу (7.4) получим

$$\begin{cases} ax_1 + b - y_1 = \varepsilon_1 \\ ax_2 + b - y_2 = \varepsilon_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ ax_n + b - y_n = \varepsilon_n \end{cases} \qquad (7.5)$$

где $\varepsilon_1, \varepsilon_2, \dots \varepsilon_n$ **- уклонения.**

По методу наименьших квадратов подберем неизвестные параметры a, b

таким образом, чтобы полученная величина

$$U = \sum_{i=1}^{n} \varepsilon_i^2 \geq 0$$

была наименьшей.

С учетом (7.5) имеем

$$U = \sum_{i=1}^{n} \varepsilon_i^2 = (ax_1 + b - y_1)^2 +$$

$$+ (ax_2 + b - y_2)^2 + \ldots + (ax_n + b - y_n)^2. \quad (7.6)$$

Переменная величина U

является функцией двух переменных a и b.

Чтобы функция U

получила возможно меньшее значение при выбранных a и b,

необходимо выполнение условия

$$\frac{\partial U}{\partial a} = 0; \frac{\partial U}{\partial b} = 0. \qquad (7.7)$$

Найдем частные производные

$$\frac{\partial U}{\partial a} = \left[(ax_1 + b - y_1)^2 + (ax_2 + b - y_2)^2 + \ldots + (ax_n + b - y_n)^2\right]'_a =$$

$$= 2(ax_1 + b - y_1)x_1 + 2(ax_2 + b - y_2)x_2 + \ldots + 2(ax_n + b - y_n)x_n =$$

$$= 2a\sum_{i=1}^{n} x_i^2 + 2b\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i y_i .$$

$$\frac{\partial U}{\partial b} = \left[(ax_1 + b - y_1)^2 + (ax_2 + b - y_2)^2 + \ldots + (ax_n + b - y_n)^2\right]'_b =$$

$$= 2(ax_1 + b - y_1) + 2(ax_2 + b - y_2) + \ldots + 2(ax_n + b - y_n) =$$

$$= 2a\sum_{i=1}^{n} x_i + 2b \cdot n - 2\sum_{i=1}^{n} y_i .$$

С учетом (7.7) получим

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i, \\ a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i. \end{cases} \qquad (7.8)$$

Решив систему определяют оптимальные значения неизвестных параметров a и b.

Замечание.

Эмпирическая формула $y = ax + b$

неплохо отражает вид зависимости между *x* и *y* в случае, когда первые разделенные разности

$$f(x_2, x_1) = \frac{y_2 - y_1}{x_2 - x_1}$$

$$f(x_3, x_2) = \frac{y_3 - y_2}{x_3 - x_2}$$

$$,\ldots, \quad f(x_n, x_{n-1}) = \frac{y_n - y_{n-1}}{x_n - x_{n-1}}.$$

мало отличаются друг от друга.

Если таблица результатов имеет постоянный шаг, то достаточно сравнивать неразделенные разности

$$\Delta y_1 = y_2 - y_1, \quad \Delta y_2 = y_3 - y_2, \quad \ldots$$

$$\Delta y_{n-1} = y_n - y_{n-1}.$$

Пример.

Результаты измерений величин *x* и *y* представлены в виде таблицы

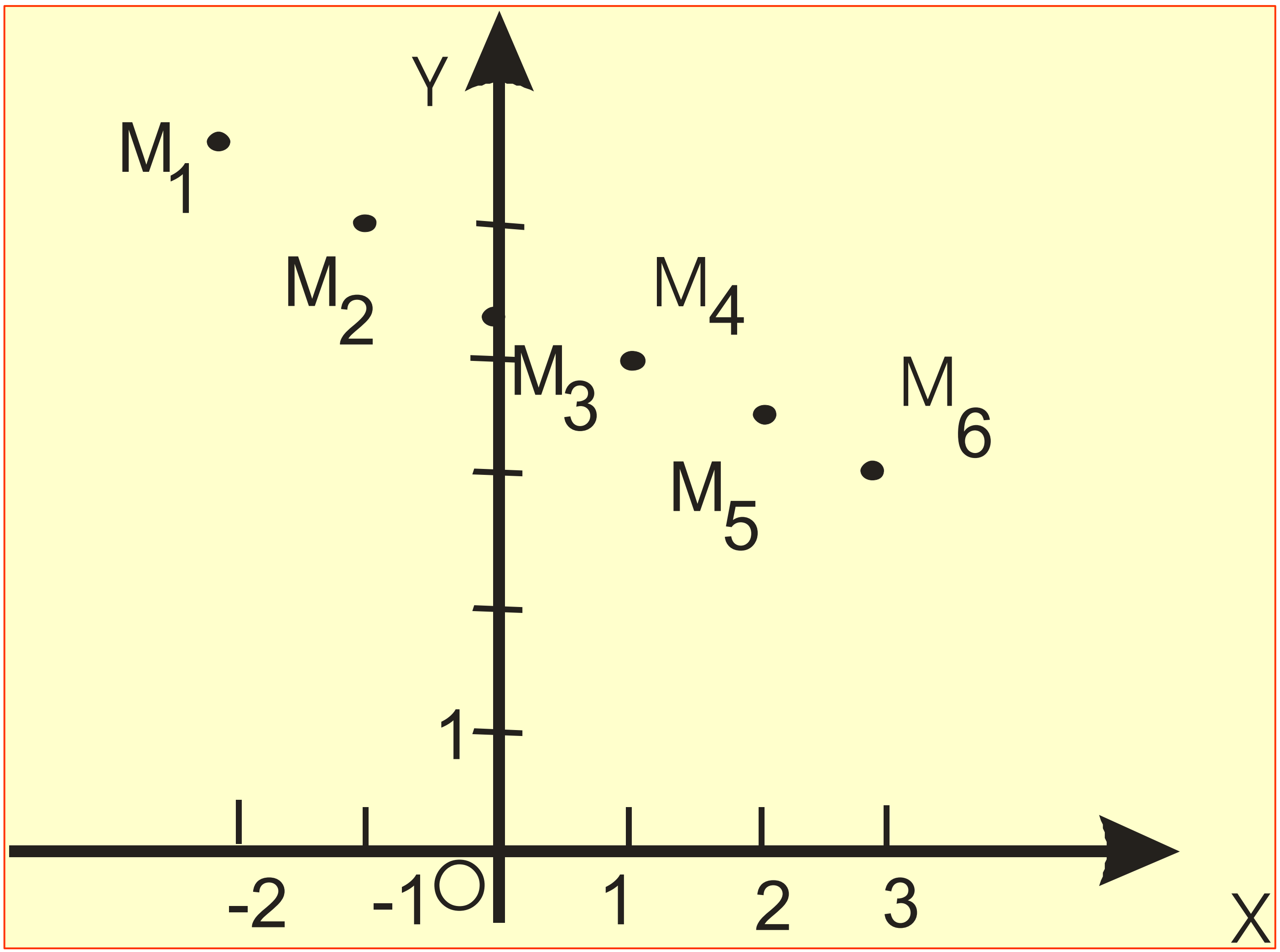
X	-2	-1	0	1	2	3
Y	5,6	5	4,3	4	3,6	3

Установить вид зависимости между этими величинами и найти параметры эмпирических формул.

Решение.

Построим в декартовой системе координат точки

$$M_1(-2;5,6), M_2(-1;5), M_3(0;4,3), M_4(1;4),$$
$$M_5(2;3,6), M_6(3,3).$$



Эти точки располагаются приблизительно на одной прямой.

Можно предположить существование между *x* и *y* линейной зависимости.

Уравнение прямой запишем в виде

$$y = ax + b.$$

Определим параметры a и b

используя систему (7.8).

Для упрощения расчетов составим таблицу

i	x_i	y_i	$x_i y_i$	x_i^2
1	-2	5,6	-11,2	4
2	-1	5	-5	1
3	0	4,3	0	0
4	1	4	4	1
5	2	3,6	7,2	4
6	3	3	9	9
Σ	3	25,5	4	19

С учетом таблицы система (7.8) примет вид

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i . \\ a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i . \end{cases}$$

$$\begin{cases} 19a + 3b = 4 \\ 3a + 6b = 25{,}5 \end{cases}$$

Решаем систему любым методом, например, Крамера

$$\Delta = \begin{vmatrix} 19 & 3 \\ 3 & 6 \end{vmatrix} = 19 \cdot 6 - 3 \cdot 3 = 105$$

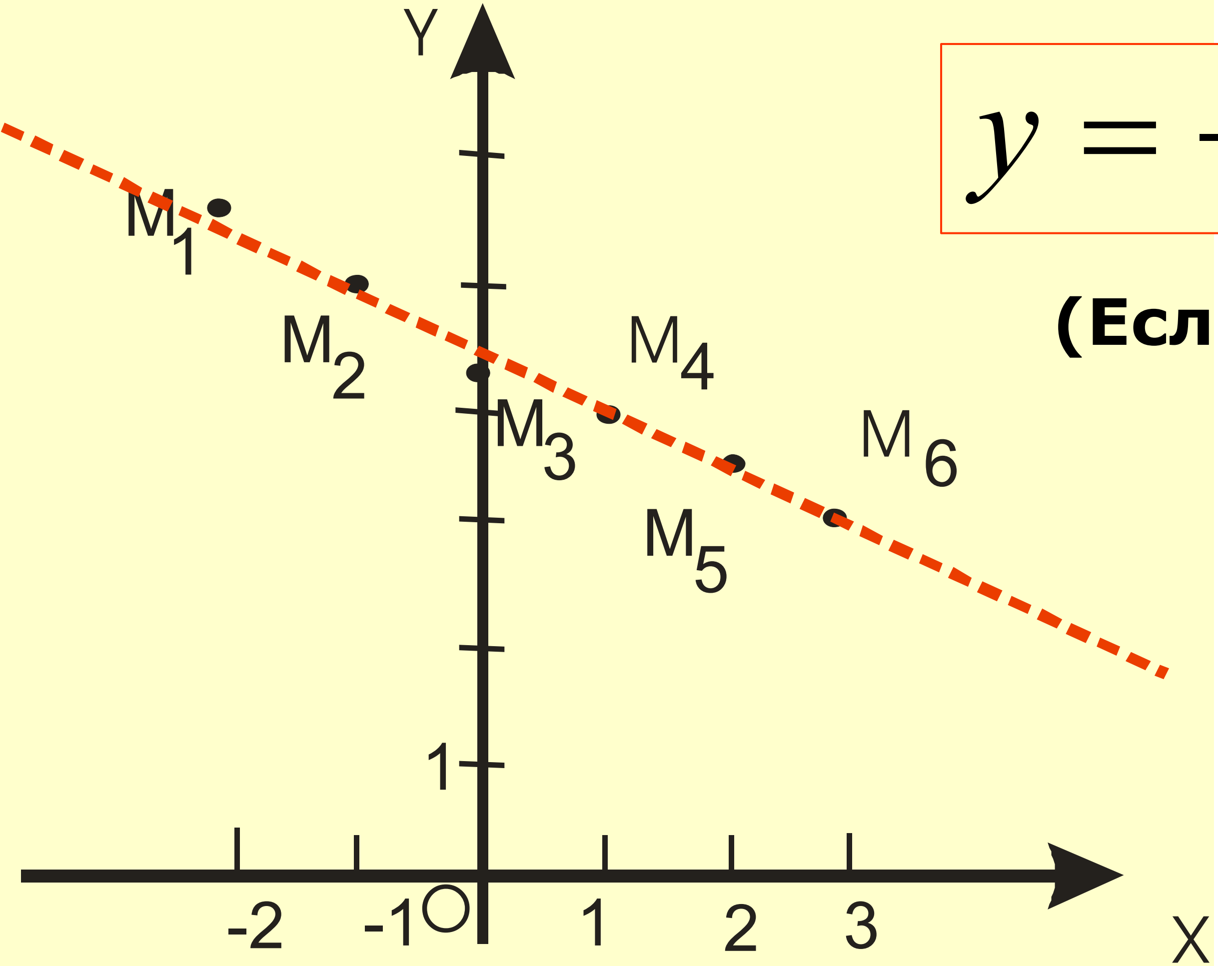
$$\Delta_a = \begin{vmatrix} 4 & 3 \\ 25{,}5 & 6 \end{vmatrix} = 4 \cdot 6 - 3 \cdot 25{,}5 = 24 - 76{,}5 = -52{,}5$$

$$\Delta_b = \begin{vmatrix} 19 & 4 \\ 3 & 25{,}5 \end{vmatrix} = 19 \cdot 25{,}5 - 3 \cdot 4 = 484{,}5 - 12 = 472{,}5$$

$$a = \frac{\Delta_a}{\Delta} = \frac{-52{,}5}{105} = -0{,}5$$

$$b = \frac{\Delta_b}{\Delta} = \frac{472{,}5}{105} = 4{,}5$$

Следовательно, зависимость между величинами *x* и *y* выражается формулой



$$y = -0{,}5x + 4{,}5.$$

(Если $x = 0,\ y = 4{,}5;$

$x = 3,\ y = 3)$

7.5 Определение параметров эмпирических формул в случае квадратичной зависимости.

Пусть в результате измерений двух зависимых величин *x* и *y* получена следующая таблица

x	x_1	x_2	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_n

Предположим, что точки

$$M_1(x_1, y_1); M_2(x_2, y_2), \ldots, M_n(x_n, y_n)$$

почти лежат на некоторой параболе.

В этом случае можно предположить, что между *x* и *y* существует квадратичная зависимость

$$y = ax^2 + bx + c \qquad (7.9)$$

Запишем выражение (7.9) в виде

$$ax^2 + bx + c - y = 0. \ (7.10)$$

Подставляя координаты точек

x_i и y_i $\quad i = 1,\ldots,n$ **в формулу (7.10) получим**

$$\begin{cases} ax_1^2 + bx_1 + c - y_1 = \varepsilon_1 \\ ax_2^2 + bx_2 + c - y_2 = \varepsilon_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ ax_n^2 + bx_n + c - y_n = \varepsilon_n \end{cases} \qquad (7.11)$$

По методу наименьших квадратов подберем неизвестные параметры таким образом, чтобы полученная величина

$$U=\sum_{i=1}^{n}\varepsilon_i^2\geq 0$$

была наименьшей.

С учетом (7.11) имеем

$$U=\sum_{i=1}^{n}\varepsilon_i^2=$$

$$=(ax_1^2+bx_1+c-y_1)^2+(ax_2^2+bx_2+c-y_2)^2+$$

$$+\ldots+(ax_n^2+bx_n+c-y_n)^2. \qquad (7.12)$$

Переменная величина U

является функцией трех переменных a, b и c.

Чтобы функция U **получила возможно меньшее значение необходимо выполнение условия**

$$\frac{\partial U}{\partial a} = 0; \frac{\partial U}{\partial b} = 0; \frac{\partial U}{\partial c} = 0. \quad (7.13)$$

Найдем частные производные

$$\frac{\partial U}{\partial a} = [(ax_1^2 + bx_1 + c - y_1)^2 + (ax_2^2 + bx_2 + c - y_2)^2 +$$

$$+ \ldots + (ax_n^2 + bx_n + c - y_n)^2]'_a =$$

$$= 2(ax_1^2 + bx_1 + c - y_1)x_1^2 + 2(ax_2^2 + bx_2 + c - y_2)x_2^2 +$$

$$+ ... + 2(ax_n^2 + bx_n + c - y_n)x_n^2 =$$

$$= 2a\sum_{i=1}^{n} x_i^4 + 2b\sum_{i=1}^{n} x_i^3 + 2c\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i^2 y_i.$$

$$\frac{\partial U}{\partial b} = [(ax_1^2 + bx_1 + c - y_1)^2 + (ax_2^2 + bx_2 + c - y_2)^2 +$$

$$+ ... + (ax_n^2 + bx_n + c - y_n)^2]'_b =$$

$$= 2(ax_1^2 + bx_1 + c - y_1)x_1 + 2(ax_2^2 + bx_2 + c - y_2)x_2 +$$

$$+ ... + 2(ax_n^2 + bx_n + c - y_n)x_n =$$

$$= 2a\sum_{i=1}^{n} x_i^3 + 2b\sum_{i=1}^{n} x_i^2 + 2c\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i y_i .$$

$$\frac{\partial U}{\partial c} = [(ax_1^2 + bx_1 + c - y_1)^2 + (ax_2^2 + bx_2 + c - y_2)^2 +$$

$$+ \ldots + (ax_n^2 + bx_n + c - y_n)^2]'_c =$$

$$= 2(ax_1^2 + bx_1 + c - y_1) + 2(ax_2^2 + bx_2 + c - y_2) +$$

$$+ \ldots + 2(ax_n^2 + bx_n + c - y_n) =$$

$$= 2a\sum_{i=1}^{n} x_i^2 + 2b\sum_{i=1}^{n} x_i + 2c \cdot n - 2\sum_{i=1}^{n} y_i .$$

С учетом (7.13) получим

$$\begin{cases} a\sum\limits_{i=1}^{n} x_i^4 + b\sum\limits_{i=1}^{n} x_i^3 + c\sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i^2 y_i. \\ a\sum\limits_{i=1}^{n} x_i^3 + b\sum\limits_{i=1}^{n} x_i^2 + c\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i y_i. \\ a\sum\limits_{i=1}^{n} x_i^2 + b\sum\limits_{i=1}^{n} x_i + c\cdot n = \sum\limits_{i=1}^{n} y_i. \end{cases} \qquad (7.14)$$

Решив систему определяют оптимальные значения неизвестных параметров a, b и $c.$

Пример.

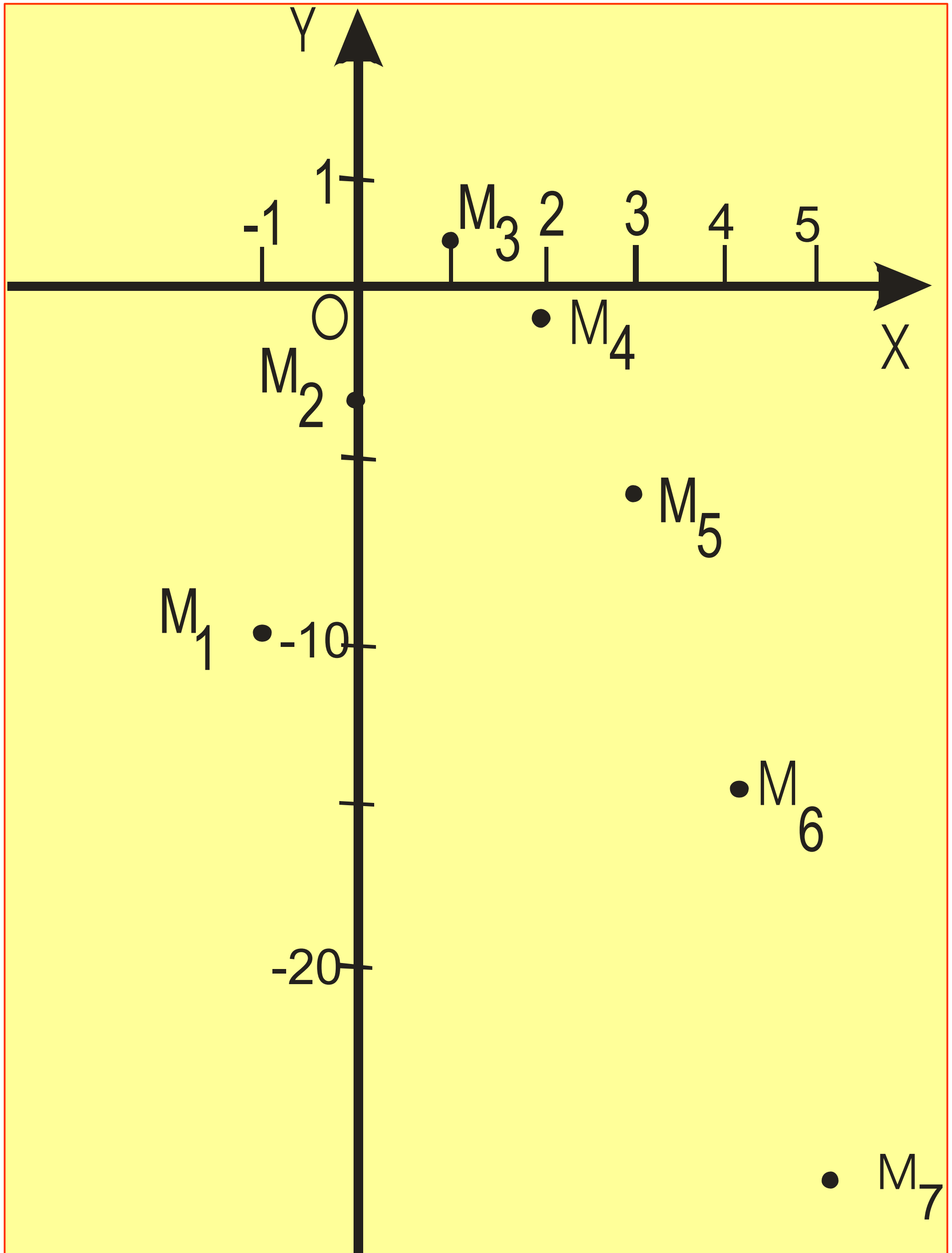
Результаты измерений величин *x* и *y* представлены в виде таблицы

X	-1	0	1	2	3	4	5
Y	-9,8	-3.1	0,3	-1,2	-6,1	-14,7	-28,2

Установить вид зависимости между этими величинами и найти параметры эмпирических формул.

Решение

Построим в декартовой системе координат точки



$M_1(-1;-9{,}8)$, $M_2(0;-3{,}1)$, $M_3(1;0{,}3)$, $M_4(2;-1{,}2)$, $M_5(3;-6{,}1)$, $M_6(4,-14{,}7)$, $M_7(5;-28{,}2)$.

Составим таблицу

i	x_i	x_i^2	x_i^3	x_i^4	y_i	$x_i y_i$	$x_i^2 y_i$
1	**-1**	**1**	**-1**	**1**	**-9,8**	**9,8**	**-9,8**
2	**0**	**0**	**0**	**0**	**-3,1**	**0**	**0**
3	**1**	**1**	**1**	**1**	**0,3**	**0,3**	**0,3**
4	**2**	**4**	**8**	**16**	**-1,2**	**-2,4**	**-4,8**
5	**3**	**9**	**27**	**81**	**-6,1**	**-18,3**	**-54,9**
6	**4**	**16**	**64**	**256**	**-14,8**	**-58,8**	**-233,2**
7	**5**	**25**	**125**	**625**	**-28,2**	**-141,0**	**-705,0**
$\sum$	**14**	**56**	**224**	**980**	**-62,8**	**-210,4**	**-1007,4**

Тогда система (7.14) принимает вид

$$\begin{cases} a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 y_i. \\ a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i. \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i + c\cdot n = \sum_{i=1}^{n} y_i. \end{cases}$$

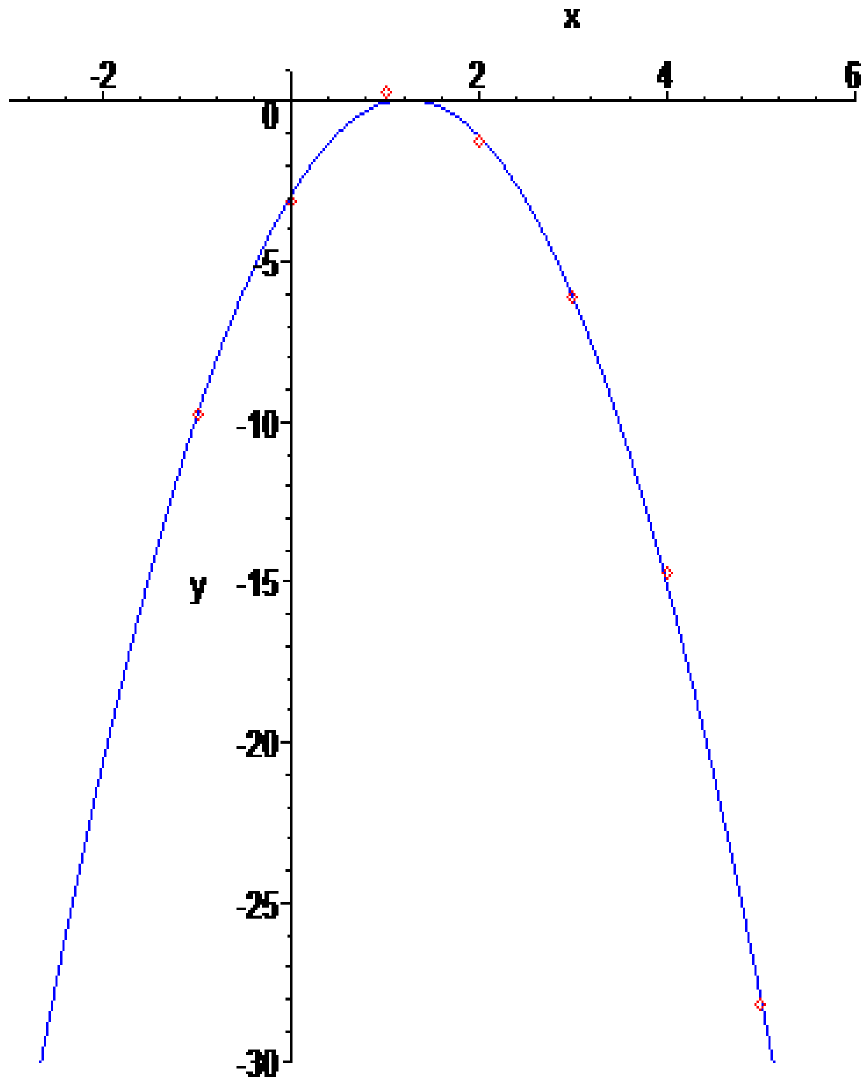
$$\begin{cases} 980a + 224b + 56c = -1007{,}4 \\ 224a + 56b + 14c = -210{,}4 \\ 56a + 14b + 7c = -62{,}8 \end{cases}$$

Решая систему получи

$$a = -1{,}974,$$

$$b = 4{,}867,$$

$$c = -2{,}914$$



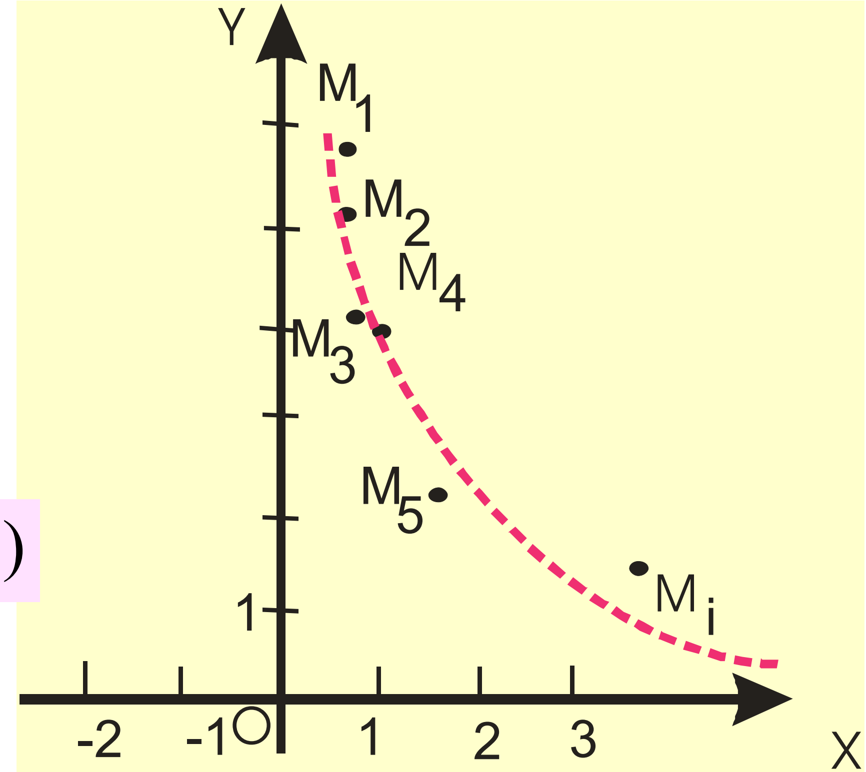
Искомая зависимость между *x* и *y* имеет вид

$$y = -1{,}974x^2 + 4{,}867x - 2{,}914.$$

7.6 Определение параметров эмпирических формул для гиперболической и показательной зависимостей.

Пусть в результате измерений двух зависимых величин *x* и *y* получена следующая таблица

x	x_1	x_2	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_n



Предположим, что точки

$$M_1(x_1, y_1); M_2(x_2, y_2), \ldots, M_n(x_n, y_n)$$

почти лежат на некоторой гиперболе

$$y = \frac{a}{x} + b, \quad (7.15)$$

где a и b **- параметры, подлежащие определению.**

Введем новую переменную $z = \frac{1}{x}$,

тогда уравнение (7.15) будет иметь вид

$$y = az + b,$$

т.е. y и z **связаны линейной зависимостью с параметрами** a и b.

В этом случае по методу наименьших квадратов параметры определяются из системы (7.8), где вместо x_i

произведена замена $z_i = \dfrac{1}{x_i}, i = 1, \ldots, n.$

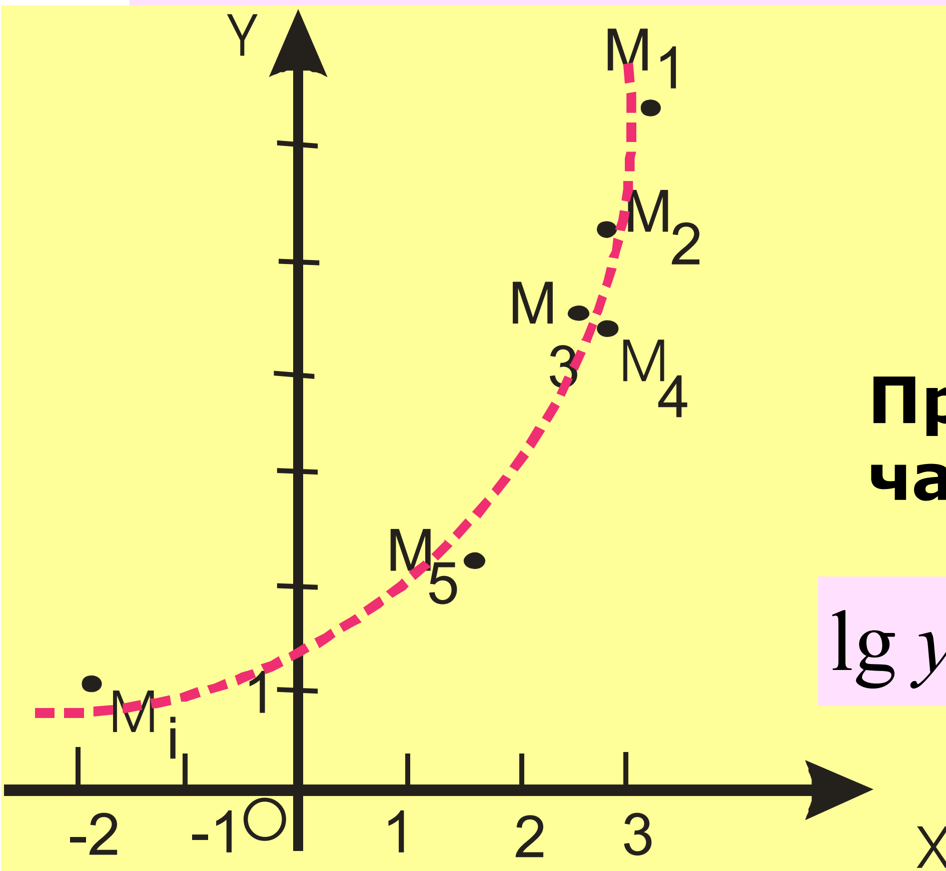
Таким образом получим систему

$$\begin{cases} a\sum_{i=1}^{n} z_i^2 + b\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} z_i y_i. \\ a\sum_{i=1}^{n} z_i + bn = \sum_{i=1}^{n} y_i. \end{cases}$$

$$\begin{cases} a\sum_{i=1}^{n} \dfrac{1}{x_i^2} + b\sum_{i=1}^{n} \dfrac{1}{x_i} = \sum_{i=1}^{n} \dfrac{1}{x_i} y_i. \\ a\sum_{i=1}^{n} \dfrac{1}{x_i} + bn = \sum_{i=1}^{n} y_i. \end{cases} \qquad (7.16)$$

Пусть анализ связи между *x* и *y* привел к выбору в качестве формы зависимости *y* от *x* показательной формы

$$y = b \cdot a^{x}, a > 0; a \neq 1. \quad (7.17)$$



Пролагорифмируем обе части уравнения (7.17)

$$\lg y = \lg(b \cdot a^{x}) = \lg b + \lg a^{x},$$

$$\lg y = x \lg a + \lg b. \quad (7.18)$$

Переменные $\lg y$ и x

связаны линейной зависимостью с параметрами

$$\lg a \text{ и } \lg b.$$

Воспользуемся системой (7.8), в которой заменяем a на $\lg a$, b на $\lg b$, y_i на $\lg y_i$.

Получаем систему

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i. \\ a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i. \end{cases}$$

$$\begin{cases} \lg a \sum_{i=1}^{n} x_i^2 + \lg b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i \lg y_i. \\ \lg a \sum_{i=1}^{n} x_i + \lg b \cdot n = \sum_{i=1}^{n} \lg y_i. \end{cases} \quad (7.19)$$

Решив систему (7.19), находим, $\lg a$ и $\lg b$,

а потом определяем параметры a и b.

Пример

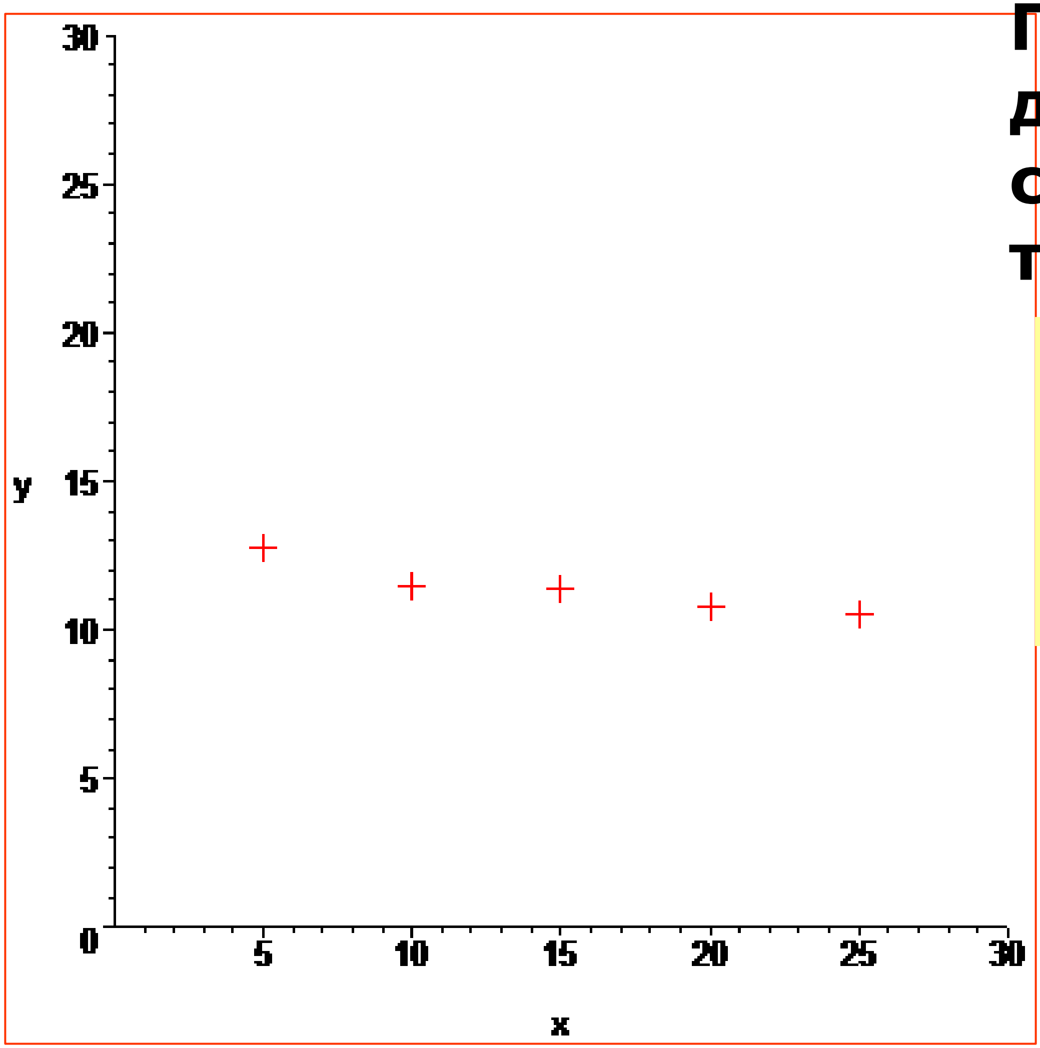
Установить форму связи между зависимыми величинами *x* и *y* и найти уравнение этой зависимости

X	5	10	15	20	25
Y	12,8	11.5	11,4	10,8	10,5

Решение

Построим в декартовой системе координат точки

$$M_1(5;12,8), M_2(10;11,5),$$
$$M_3(15;11,4),$$
$$M_4(20;10,8), M_5(25;10,5).$$



Из рисунка видно, что между переменными существует гиперболическая зависимость

$$y = \frac{a}{x} + b.$$

i	x_i	$\frac{1}{x_i}$	$\frac{1}{x_i^2}$	y_i	$\frac{y_i}{x_i}$
1	5	0,2	0,04	12,8	2,56
2	10	0,1	0,01	11,5	1,15
3	15	0,067	0,0045	11,4	0,76
4	20	0,05	0,0025	10,8	0,54
5	25	0,04	0,0016	10,5	0,42
$\sum$	75	0,457	0,0586	57,0	5,43

Результаты расчетов приведем в таблице

Система уравнений (7.16)

$$\begin{cases} a\sum\limits_{i=1}^{n}\dfrac{1}{x_i^2}+b\sum\limits_{i=1}^{n}\dfrac{1}{x_i}=\sum\limits_{i=1}^{n}\dfrac{1}{x_i}y_i. \\ a\sum\limits_{i=1}^{n}\dfrac{1}{x_i}+bn=\sum\limits_{i=1}^{n}y_i. \end{cases}$$

примет вид

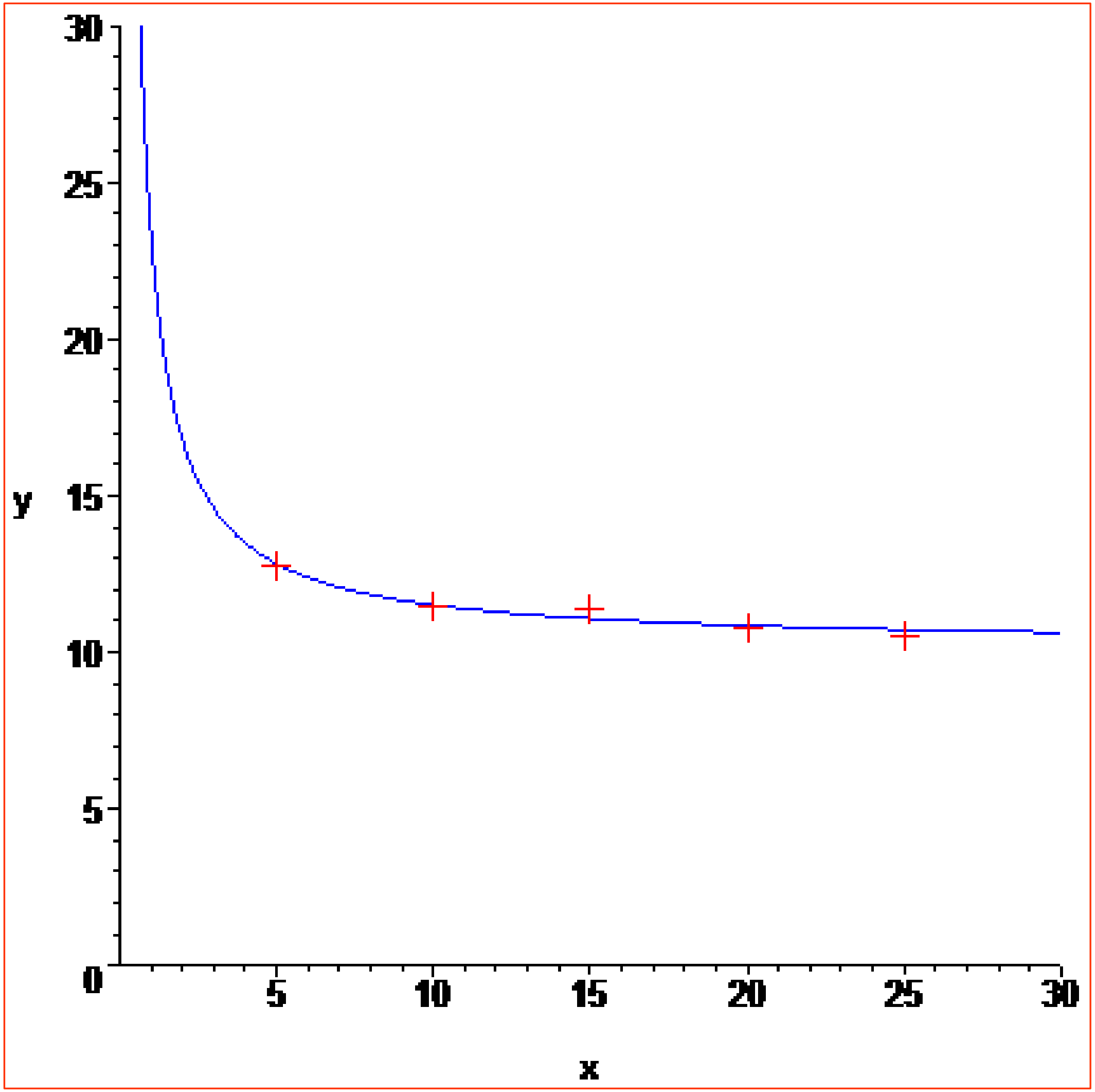
$$\begin{cases} 0{,}0586a+0{,}457b=5{,}43 \\ 0{,}457a+5b=57. \end{cases}$$

Решая ее, найдем

$$a=13{,}113;\quad b=10{,}2$$

и тогда уравнение примет вид

$$\boxed{y=\frac{13{,}113}{x}+10{,}2.}$$


30
25
20
15
10
5
0
y
5
10
15
20
25
30
x

7.7 Определение параметров эмпирических формул в случае сгруппированных данных

Пусть в результате некоторого эксперимента было получено довольно большое число данных, среди которых есть повторяющиеся.

Пусть одно и тоже значение x_i встретилось n_{x_i} раз, а одно и то же значение y_j встретилось n_{y_j} раз,

одна и та же пара значений (x_i, y_j) встретилась n_{ij} раз, $i = 1,.., s; \quad j = 1,..., t.$

Поэтому данные наблюдений можно сгруппировать в виде таблицы, которую называют корреляционной.

$X \backslash Y$	y_1	y_2	$\cdots$	y_t	n_{x_i}
x_1	$n_{x_1 y_1}$	n_{x1y2}	$\cdots$	n_{x1yt}	n_{x1}
x_2	n_{x2y1}	n_{x2y2}	$\cdots$	n_{x2yt}	n_{x2}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_s	$n_{x_s y_1}$	n_{xsy2}	$\cdots$	n_{xsyt}	n_{xs}
n_{y_j}	n_{y_1}	n_{y_2}	$\cdots$	n_{y_t}	

Объем выборки можно подсчитать тремя способами:

$$\sum_{i=1}^{s}\sum_{j=1}^{t} n_{ij} = n;$$

$$\sum_{i=1}^{s} n_{x_i} = n$$

$$\sum_{j=1}^{t} n_{y_j} = n$$

По корреляционной таблице подсчитывают

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{s} x_i n_{x_i}$$

$$\sum_{i=1}^{n} y_i = \sum_{j=1}^{t} y_j n_{y_j}$$

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{s}\sum_{j=1}^{t} x_i y_j n_{ij}$$

и подставив полученные значения в системы уравнений (7.8), (7.14), (7.16) и (7.19) для различных видов зависимостей, получим системы уравнений.

Например, для линейной зависимости система принимает вид

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i. \\ a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i. \end{cases}$$

$$\begin{cases} a\sum_{i=1}^{s} x_i^2 n_{x_i} + b\sum_{i=1}^{n} x_i n_{x_i} = \sum_{i=1}^{s}\sum_{j=1}^{t} x_i y_j n_{ij}. \\ a\sum_{i=1}^{s} x_i n_{x_i} + bn = \sum_{i=1}^{t} y_j n_{y_j}. \end{cases} \qquad (7.20)$$

Замечание.

Под групповыми средними понимают

$$\overline{y_i} = \frac{\sum\limits_{j=1}^{t} y_j n_{ij}}{n_{x_i}} \qquad \overline{x_j} = \frac{\sum\limits_{i=1}^{s} x_i n_{ij}}{n_{y_j}}$$

Если необходимо определить вид зависимости

$y = f(x),$ **то отмечают точки с координатами**

$$(x_i, \overline{y_i}), \quad i = 1, \ldots, s.$$

И по их группировке вдоль некоторой линии определяют характер зависимости.

Аналогично и для зависимости $x = \varphi(y).$

Разделим члены обоих уравнений системы (7.20) на объем совокупности *n*, получим

$$\begin{cases} a\frac{1}{n}\sum_{i=1}^{s} x_i^2 n_{x_i} + b\frac{1}{n}\sum_{i=1}^{n} x_i n_{x_i} = \frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{t} x_i y_j n_{ij}. \\ a\frac{1}{n}\sum_{i=1}^{s} x_i n_{x_i} + b = \frac{1}{n}\sum_{i=1}^{t} y_j n_{y_j}. \end{cases}$$

Обозначив средние арифметические значения переменных *x и* у за $\overline{x}$ и $\overline{y},$

а среднюю арифметическую их произведений за $\overline{xy},$ получим следующую систему

$$\begin{cases} a\overline{x^2} + b\bar{x} = \overline{xy} & (7.21) \\ a\bar{x} + b = \bar{y} \end{cases}$$

Из второго уравнения выразим *b*

$$b = \bar{y} - a\bar{x}.$$

Подставим его в уравнение

$$y = ax + b,$$

получим

$$y = ax + \bar{y} - a\bar{x},$$

$$y - \bar{y} = a(x - \bar{x}). \quad (7.22)$$

Выражение (7.22) называют уравнением прямой линии регрессии, которая проходит через точку с координатами $(\overline{x}, \overline{y})$.

При этом коэффициент a **называют коэффициентом регрессии *y* на *x* и обозначают символом** $\rho_{y/x}$,

т.е. $a = \rho_{y/x}$.

Подставив $b = \overline{y} - \rho_{y/x}\overline{x}$ **в первое уравнение системы (7.21) получим**

$$y = ax + \overline{y} - a\overline{x},$$ или $$y - \overline{y} = a(x - \overline{x}). \quad (7.22)$$

Выражение (7.22) называют уравнением прямой линии регрессии, которая проходит через точку с координатами $(\overline{x}, \overline{y})$.

При этом коэффициент a **называют коэффициентом регрессии *y* на *x* и обозначают символом** $\rho_{y/x}$, $a = \rho_{y/x}$.

Подставив $b = \overline{y} - a\overline{x}$ **в первое уравнение системы (7.21) получим**

$$\rho_{y/x}\,\overline{x^2}+(\overline{y}-\rho_{y/x}\,\overline{x})\,\overline{x}=\overline{xy}$$

$$\rho_{y/x}\,\overline{x^2}+\overline{x}\cdot\overline{y}-\rho_{y/x}(\overline{x})^2=\overline{xy},$$

Откуда

$$\rho_{y/x}=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\overline{x^2}-(\overline{x})^2}\qquad(7.23)$$

Чтобы найти уравнение прямой линии регрессии(7.22), достаточно определить

$\rho_{y/x}$, **по формуле (7.23)** $\overline{x}, \overline{y}.$

7.8 Методы расчета свободных характеристик. Условные варианты

Равноотстоящими называют варианты, которые образуют арифметическую прогрессию с разностью *h*.

Условными называют варианты, определяемые равенством

$$u_i = \frac{x_i - C}{h},$$

где *C*- ложный нуль (новое начало отсчета), *h* – шаг, т.е. разность между любыми двумя соседними первоначальными вариантами.

Замечание. **В качестве ложного нуля можно принять любую варианту.**

Вычисления будут наиболее простыми, если выбрать в качестве ложного нуля варианту, которая расположена примерно в середине вариационного ряда.

Часто такая варианта имеет наибольшую частоту

Пример.

Найти условные варианты статистического распределения

x_i	23,6	28,6	33,6	38,6	43,6
n_i	5	20	50	15	10

Решение.

В качестве *C* выберем варианту 33,6.

$$h = 28{,}6 - 23{,}6 = 33{,}6 - 28{,}6 = \ldots = 5$$

$$u_1 = \frac{x_1 - C}{h} = \frac{23{,}6 - 33{,}6}{5} = -2.$$

Аналогично получим

$$u_2 = \frac{x_2 - C}{h} = \frac{28{,}6 - 33{,}6}{5} = -1$$

$$u_3 = 0,\ u_4 = 1,\ u_5 = 2.$$

Видно, что условные варианты – небольшие числа. Поэтому с ними работать проще, чем с первоначальными вариантами.

Если первоначальные варианты не являются равноотстоящими, то переходят к вариантам

$$u_i = x_i - C. \quad (7.24)$$

В этом случае

$$\overline{x}_B = C + \frac{\sum_i n_i u_i}{n}. \quad (7.25)$$

Действительно, подставим (7.24) в (7.25)

$$\overline{x}_B = C + \frac{\sum_i n_i (x_i - C)}{n} = C + \frac{\sum_i n_i x_i - C\sum_i n_i}{n} =$$

$$= C + \frac{\sum_i n_i x_i - Cn}{n} = \frac{\sum_i n_i x_i}{n}$$

Пример.

Найти выборочную среднюю по данному распределению выборки

x_i	1250	1270	1280
n_i	2	5	3

Решение.

Первоначальные варианты большие числа поэтому перейдем к условным вариантам

$$u_i = x_i - C, \quad \mathrm{C} = 1270$$

$$u_i = x_i - 1270.$$

Получим распределение условных вариант

x_i	-20	0	10
n_i	2	5	3

Найдем искомую выборочную среднюю

$$\overline{x}_B = C + \frac{\sum\limits_i n_i u_i}{n} = 1270 + \frac{2(-20) + 0\cdot 5 + 10\cdot 3}{10} =$$

$$= 1270 - 1 = 1269.$$

Выборочные дисперсии являются смещенной оценкой генеральной дисперсии и определяются, при переходе к условным вариантам, по формуле

$$D_B(X) = D_B(u) = \overline{u^2} - \overline{u}^2 = \frac{\sum\limits_i n_i u_i^2}{n} - \left[\frac{\sum\limits_i n_i u_i}{n}\right]^2 \quad (7.26)$$

Пример.

Найти выборочную дисперсию по данному распределению выборки объема $n = 10.$

x_i	186	192	194
n_i	2	5	3

Решение.

Перейдем к условным вариантам

$$u_i = x_i - 192.$$

Получим распределение условных вариант

u_i	-6	0	2
n_i	2	5	3

Выборочная дисперсия определяется по формуле (7.26)

$$D_B(X) = D_B(u) = \frac{\sum_i n_i u_i^2}{n} - \left[\frac{\sum_i n_i u_i}{n}\right]^2 =$$

$$= \frac{2\cdot(-6)^2 + 0 + 3\cdot 2^2}{10} - \left[\frac{2(-6) + 5\cdot 0 + 3\cdot 2}{10}\right]^2 =$$

$$= \frac{2\cdot 36 + 3\cdot 4}{10} + \left[\frac{-6}{10}\right]^2 = 8{,}04.$$

Если первоначальные варианты являются десятичными дробями с *k* десятичными знаками после запятой, то чтобы избежать действий с дробями умножают первоначальные варианты на число C^k,

т.е. переходят к условным вариантам $u_i = C^k x_i.$

При этом дисперсия увеличивается в C^k **раз**

$$D_B(x) = \frac{D_B(u)}{C^k}. \quad (7.27)$$

Пример.

Найти выборочную дисперсию по данному распределению выборки объема $n = 10.$

x_i	0,01	0,04	0,08
n_i	5	3	2

Решение.

Перейдем к условным вариантам $u_i = 100x_i$.

Получим распределение условных вариант

u_i	1	4	8
n_i	5	3	2

$$D_B(u) = \frac{\sum_i n_i u_i^2}{n} - \left[\frac{\sum_i n_i u_i}{n}\right]^2 =$$

$$= \frac{5\cdot 1 + 3\cdot 16 + 2\cdot 64}{10} - \frac{5\cdot 1 + 3\cdot 4 + 2\cdot 8}{10} =$$

$$= 18{,}1 - 3{,}3^2 = 7{,}21$$

$$D_B(X) = \frac{D_B(u)}{100^2} = 0{,}0007.$$