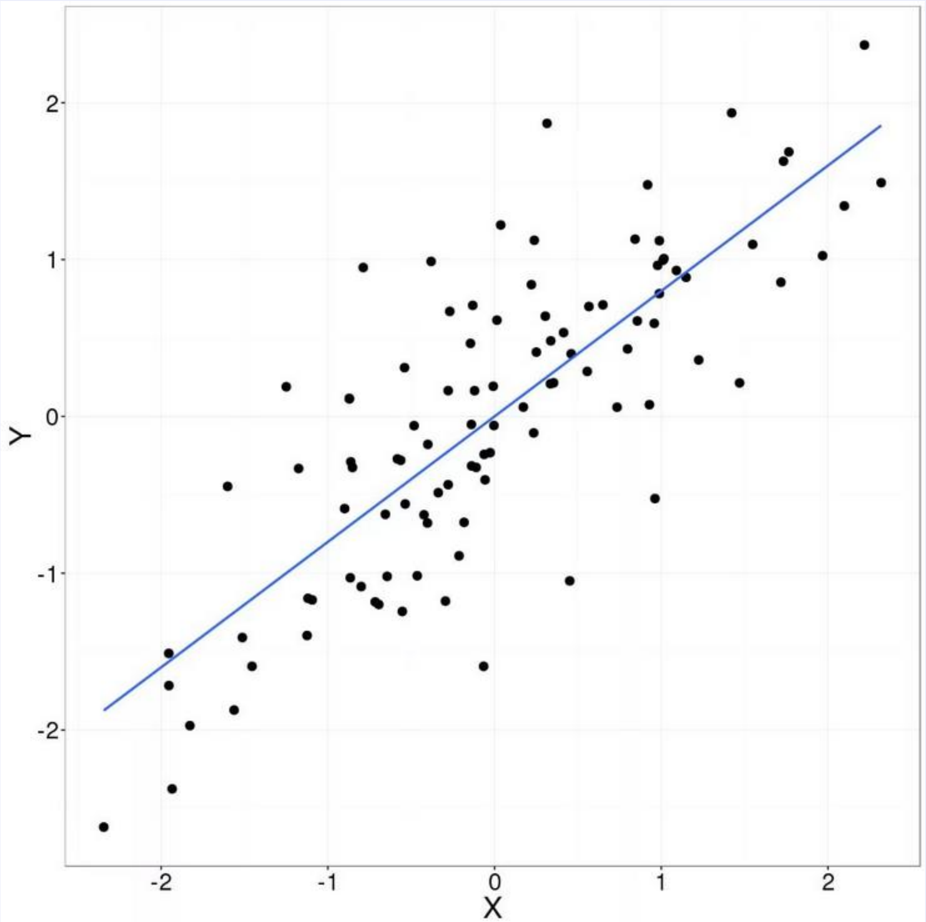
50 ЛЕТ
ИНСТИТУТ
ЭКОНОМИКИ
УрО РАН

Модифицированный метод градиентного покоординатного спуска по узловым прямым

Докладчик: Голованов О.А.

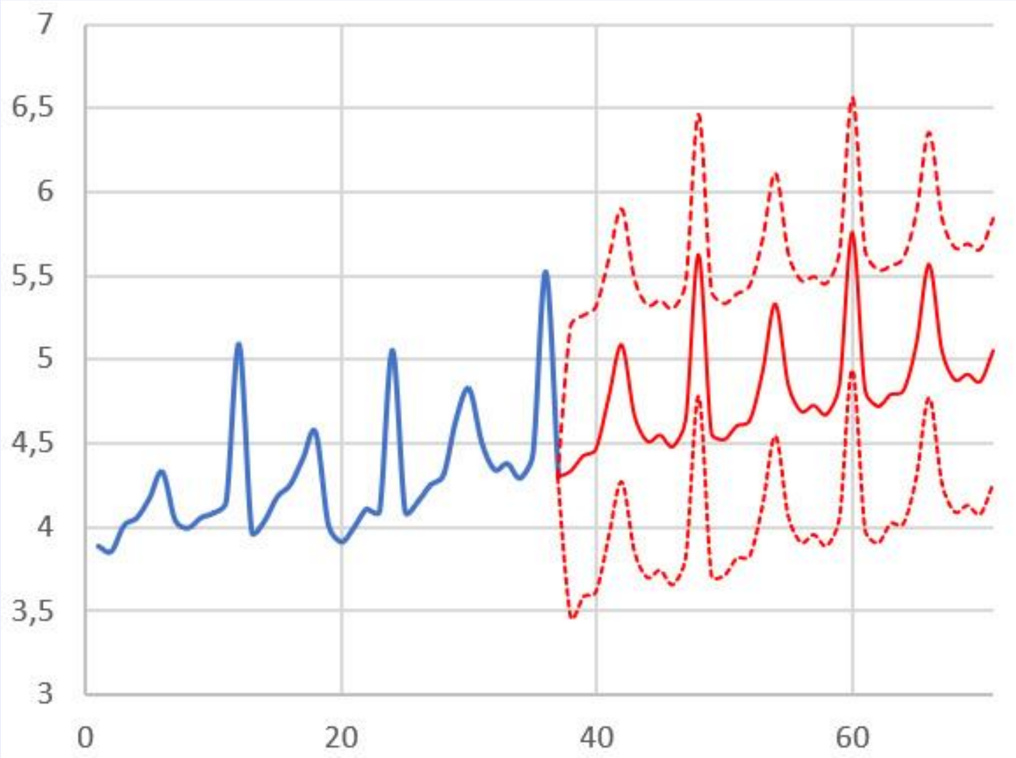
Линейная регрессия



Y = Xa + ε

где **y** – вектор зависимых переменных, **X** – матрица объясняющих переменных, **ε** – вектор случайных отклонений, **a** – вектор искомых коэффициентов

Область применения



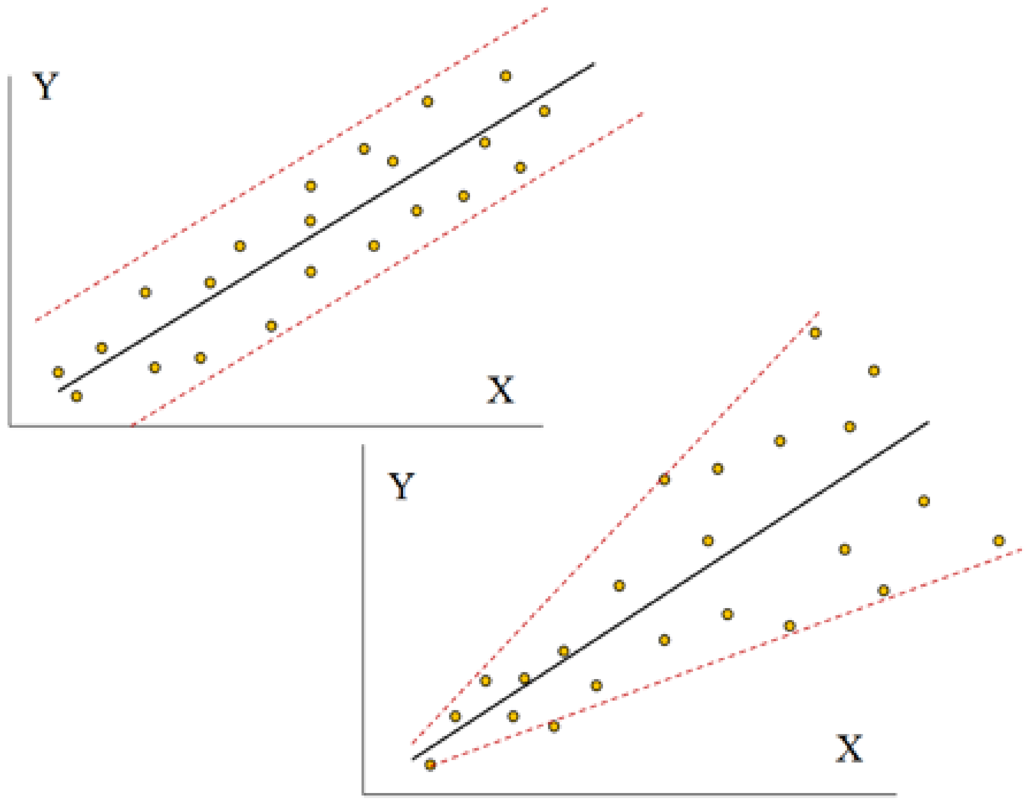
Пример прогнозирования
сред. заработной платы[1]

Модель может использоваться в сфере:

- Экономики
- Финансового анализа
- Маркетинга
- Медицинских исследований
- Социологических исследований

[5] Голованов О.А., Тырсин А.Н. Регрессионный анализ данных на основе метода наименьших модулей в динамических задачах оценивания / Заводская лаборатория. Диагностика материалов. 2023. Т. 89. № 5. С. 71–80. DOI: 10.26896/1028-6861-2023-89-5-71-80

Метод наименьших квадратов

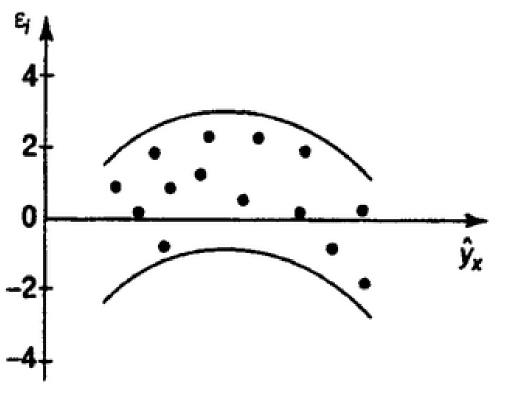


Гомоскедастичность и гетероскедастичность

$$Q(\mathbf{a}) = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{m} a_j x_{ij}\right)^2 \to \min_{\mathbf{a}\epsilon\mathbf{R}^m},$$

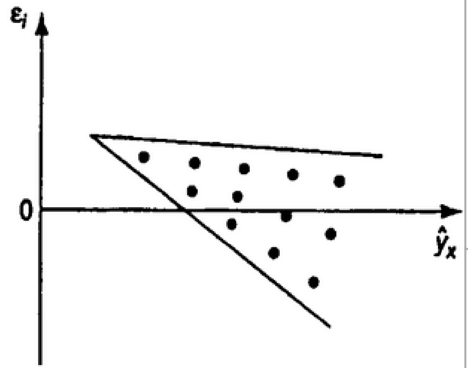
Условия Гаусса-Маркова

1. Регрессионная модель должна быть корректно ***специфицирована***, т.е. должна быть линейной относительно параметров и иметь аддитивный случайный член
2. Случайные отклонения должны:
   a) Иметь ***нулевое*** математическое ожидание;
   b) ***Постоянную*** дисперсию;
   c) Быть ***независимыми*** друг от друга;
   d) Иметь ***независимость*** от объясняющей переменной.



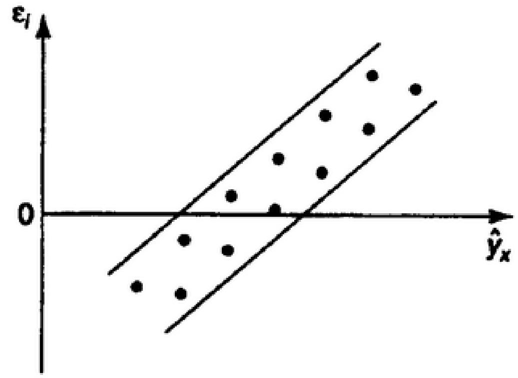
а

остатки ε_i не случайны



б

остатки ε_i не имеют постоянной дисперсии



в

остатки ε_i носят систематический характер

Условия Гаусса-Маркова

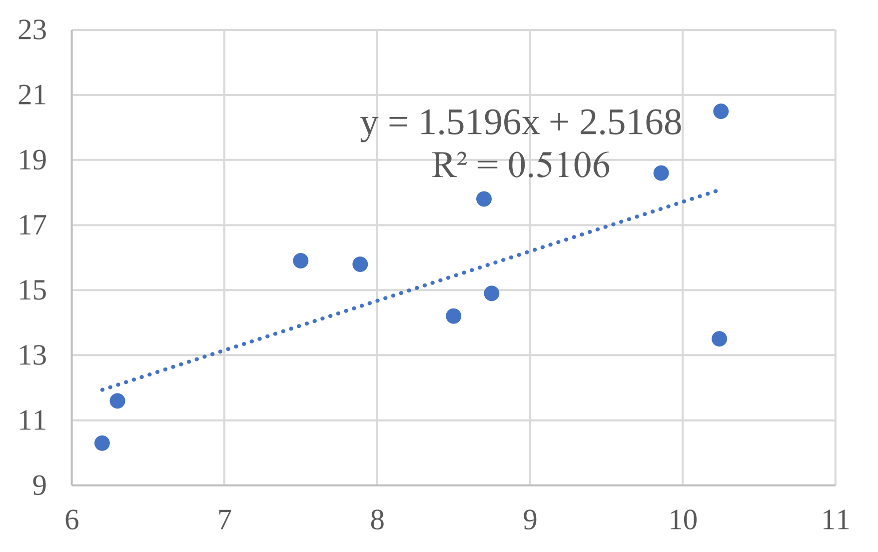
Наряду с выполнимостью указанных предпосылок при построении линейных регрессионных моделей обычно делаются еще ***некоторые предположения***, а именно:

- случайное отклонение имеет нормальный закон распределения;
- число наблюдений существенно больше числа объясняющих переменных;
- отсутствуют ошибки спецификации;
- отсутствует линейная взаимосвязь между двумя или несколькими объясняющими переменными.

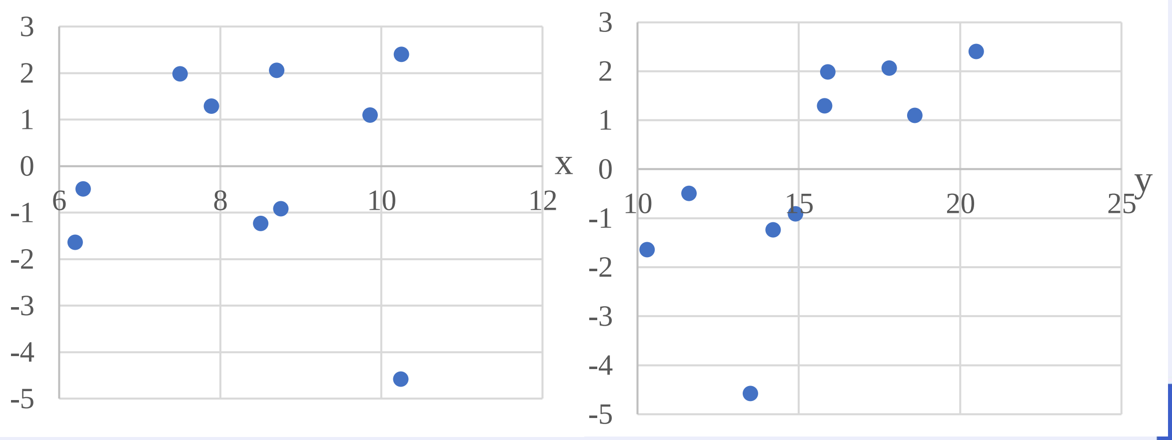
Пример

Показатели среднемесячной заработной платы под влиянием прожиточного минимума по регионам РФ за 2016 г., тыс. руб.

№	Среднедушевой прожиточный минимум на одного работающего (x)	Среднемесячная заработная плата (y)
1	8.7	17.8
2	6.3	11.6
3	7.89	15.8
4	10.24	13.5
5	10.25	20.5
6	7.5	15.9
7	8.75	14.9
8	6.2	10.3
9	9.86	18.6
10	8.5	14.2



y = 1.5196x + 2.5168



Метод наименьших модулей

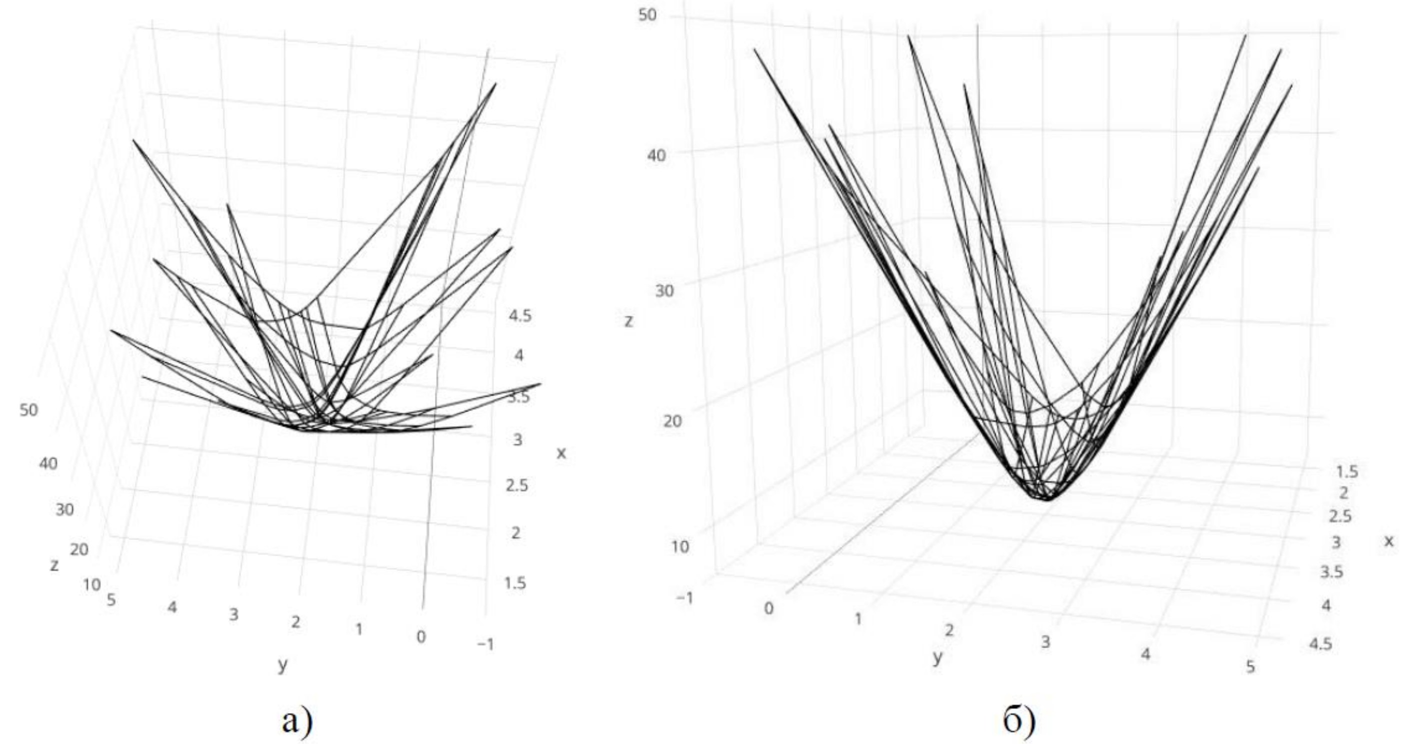
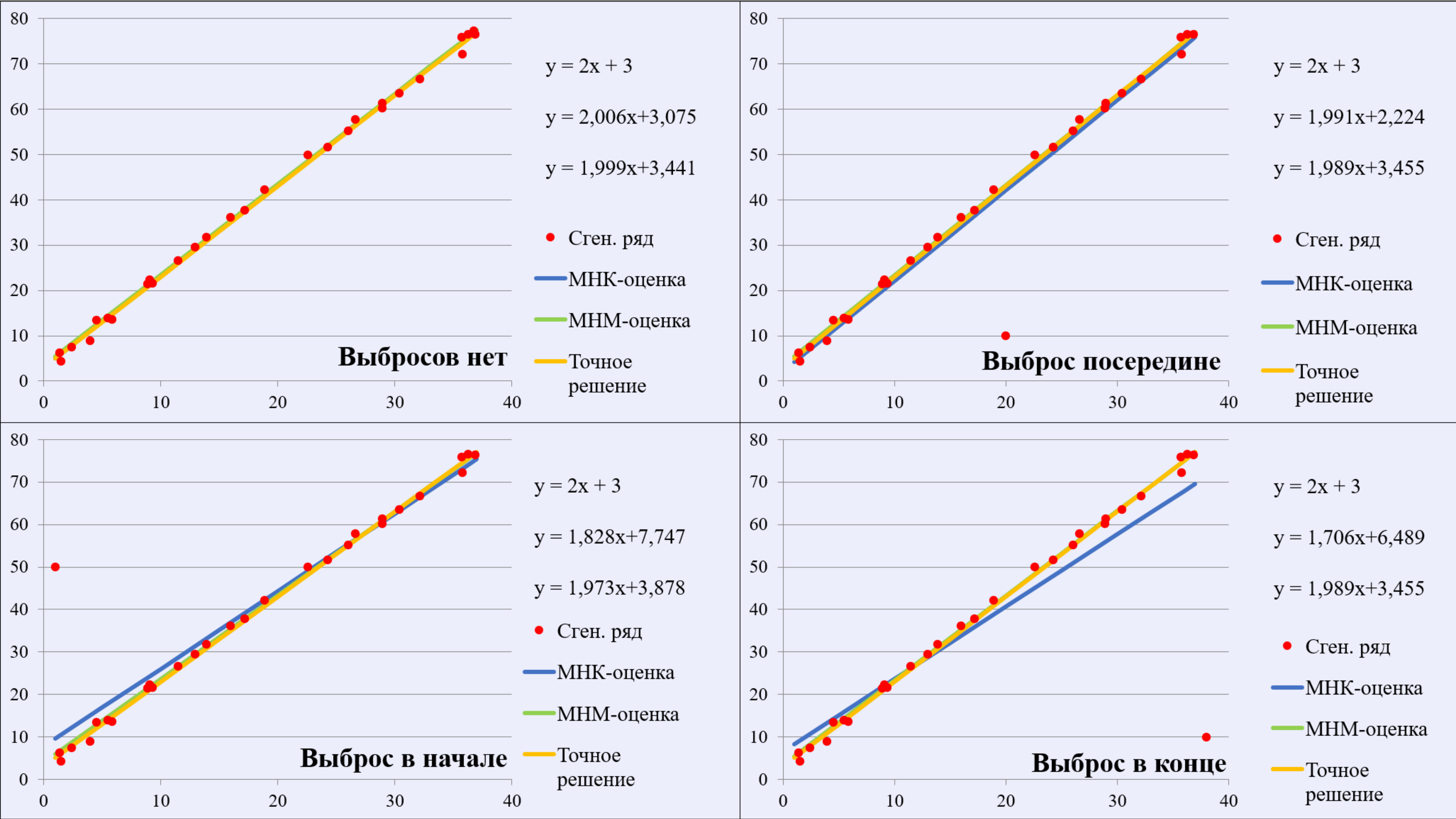


График функции Q(a) для m=2 и n=28: а) вид сверху, б) вид сбоку

$$Q(\mathbf{a}) = \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{m} a_j x_{ij} \right| \to \min_{\mathbf{a} \in \mathbf{R}^m}.$$

[1] Болдин М.В., Симонова Г.И., Тюрин Ю.Н. (1997). Знаковый статистический анализ линейных моделей. М.: Наука, Физматлит. 288 с.
[2] Basset G.W., Koenker R. (1978). Asymptotic theory of least absolute error. *Journal of the American Statistical Association*, 73, 618–622.
[3] Pollard D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometrics Theory*, 7, 186–199.


y = 2x + 3
y = 2,006x+3,075
y = 1,999x+3,441
Сген. ряд
МНК-оценка
МНМ-оценка
Точное решение
Выбросов нет
y = 2x + 3
y = 1,991x+2,224
y = 1,989x+3,455
Сген. ряд
МНК-оценка
МНМ-оценка
Точное решение
Выброс посередине
y = 2x + 3
y = 1,828x+7,747
y = 1,973x+3,878
Сген. ряд
МНК-оценка
МНМ-оценка
Точное решение
Выброс в начале
y = 2x + 3
y = 1,706x+6,489
y = 1,989x+3,455
Сген. ряд
МНК-оценка
МНМ-оценка
Точное решение
Выброс в конце

Покоординатный спуск

Пусть $\boldsymbol{\Omega}: \{\Omega_1, \Omega_2, \dots, \Omega_n\}$ будет являться множеством всех гиперплоскостей вида[4]:

$$\Omega_{\boldsymbol{i}} = \Omega(\mathbf{a}, \mathbf{x}_{\boldsymbol{i}}, y_{\boldsymbol{i}}) = y_{\boldsymbol{i}} - \langle \mathbf{a}, \mathbf{x}_{\boldsymbol{i}} \rangle = 0, (i = 1, \dots, n).$$

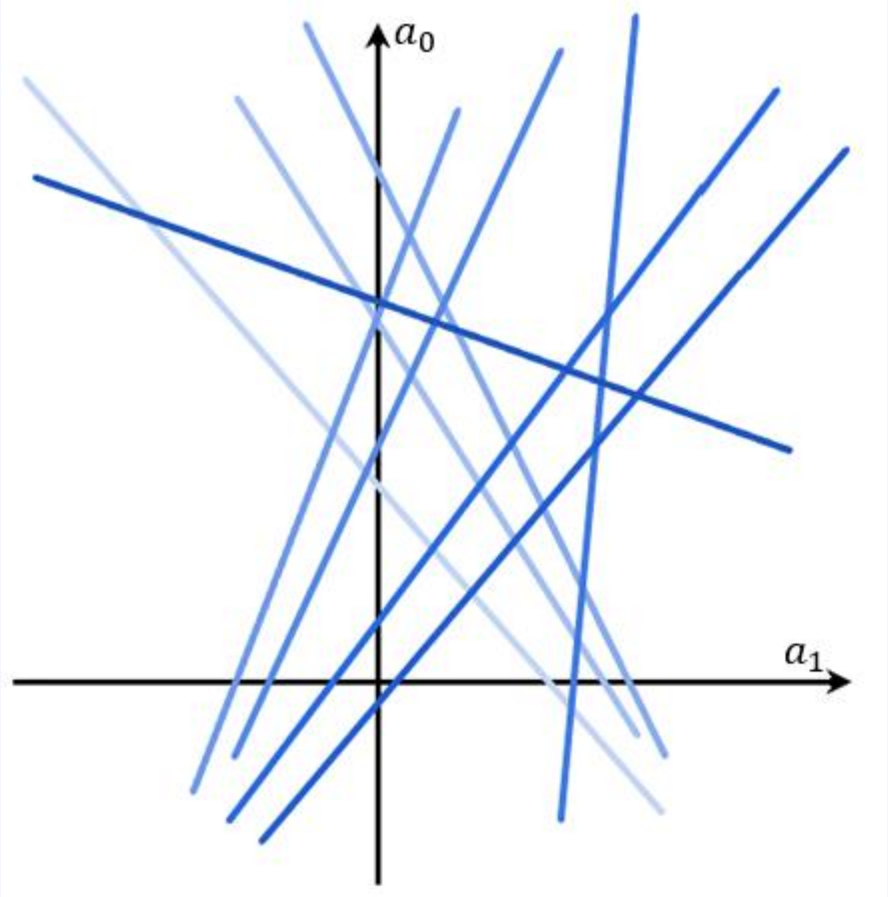
Тогда, узловая точка будет представлять собой точку пересечения m независимых гиперплоскостей

$$\mathbf{u} = \bigcap_{s \in M} \Omega_s, M = \{k_1, \dots, k_m\},$$

$$k_1 < k_2 < \dots < k_m, k_l \in \{1, \dots, n\}.$$

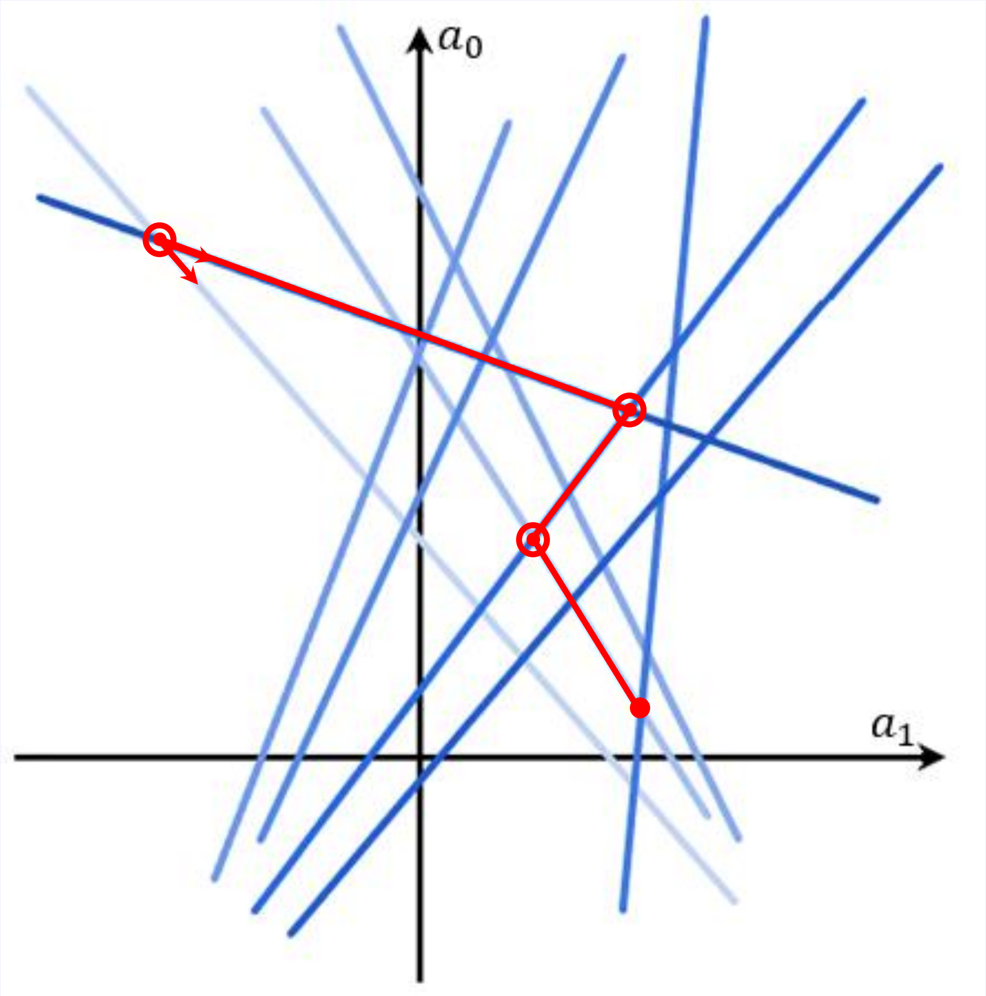
Назовем узловой такую прямую, которая будет являться пересечением независимых гиперплоскостей

$$l_{(k_1, \dots, k_{m-1})}: \bigcap \Omega_i, \ i \in \{k_1, \dots, k_{m-1}\}, k_l \in \{1, \dots, n\}.$$



$$a_0 = y_i + a_1 x_i$$

[4] Тырсин А.Н. Алгоритмы спуска по узловым прямым в задаче оценивания регрессионных уравнений методом наименьших модулей // Заводская лаборатория. Диагностика материалов. 2021. Т. 87. № 5. С. 68-75. DOI: 10.26896/1028-6861-2021-87-5-68-75



Спуск по узловым прямым

Градиентный покоординатный спуск

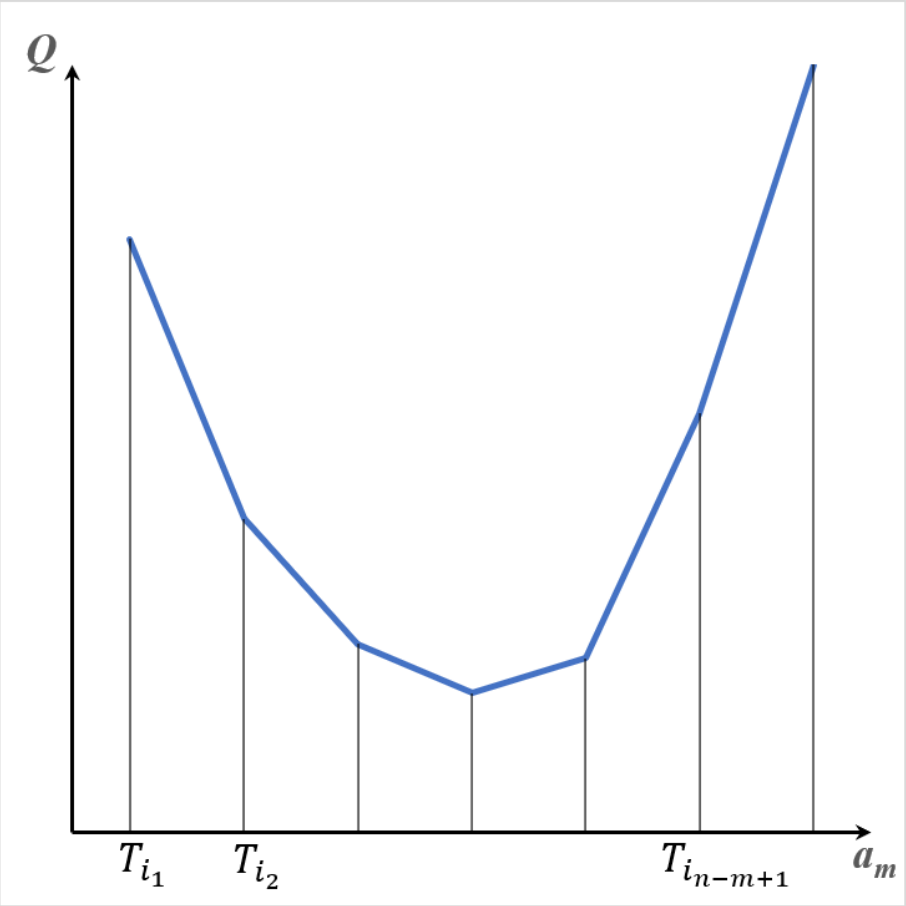
Пусть $\mathbf{u}^{(*)} = \left(u_1^{(*)}, \dots, u_m^{(*)}\right)$ - исходная узловая точка[5], тогда

$$\frac{\partial Q\left(\mathbf{u}^{(*)}\right)}{\partial \mathbf{l}_{(k_1,\dots,k_{m-1})}} = \sum_{i=1}^{n}(c_1 + x_{i2}c_2 + \dots + x_{im}c_m) \cdot \operatorname{sign}\left(\sum_{j=1}^{m} u_j^{(*)} x_{ij} - y_i\right),$$

где $\mathbf{l}_{(k_1,\dots,k_{m-1})} = (c_1, c_2, \dots, c_m)$– направляющий вектор узловой прямой $l_{(k_1,\dots,k_{m-1})}$. Если производная по направлению слева $\left.\frac{\partial Q\left(\mathbf{u}^{(*)}\right)}{\partial \mathbf{l}_{(k_1,\dots,k_{m-1})}}\right|_{\mathbf{u}_-^{(*)}}$ и справа $\left.\frac{\partial Q\left(\mathbf{u}^{(*)}\right)}{\partial \mathbf{l}_{(k_1,\dots,k_{m-1})}}\right|_{\mathbf{u}_+^{(*)}}$ меняет знак, то в исследуемой точке находится экстремум и целевая функция на прямой достигает своего минимального значения.

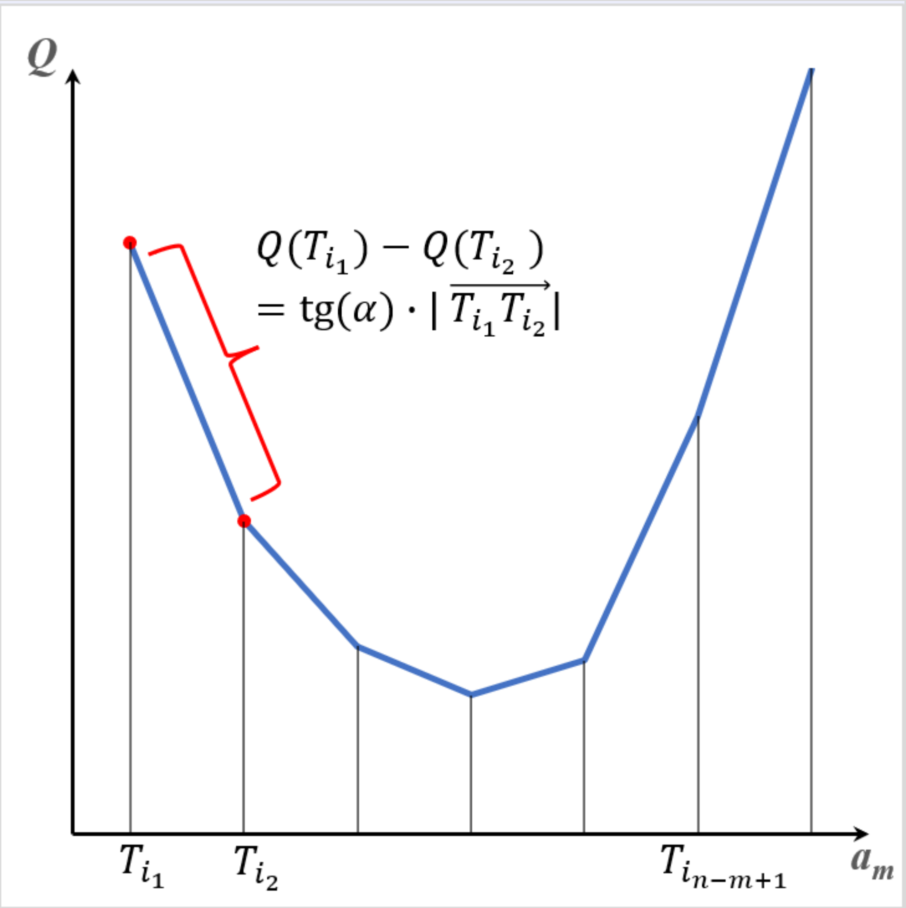
[5] Голованов О.А., Тырсин, А. Н. Динамическое регрессионное моделирование на основе градиентного спуска по узловым прямым / А. Н. Тырсин, О. А. Голованов // Современные наукоемкие технологии. – 2021. – № 10. – С. 88-93. – DOI 10.17513/snt.38859.

Учет угла наклона



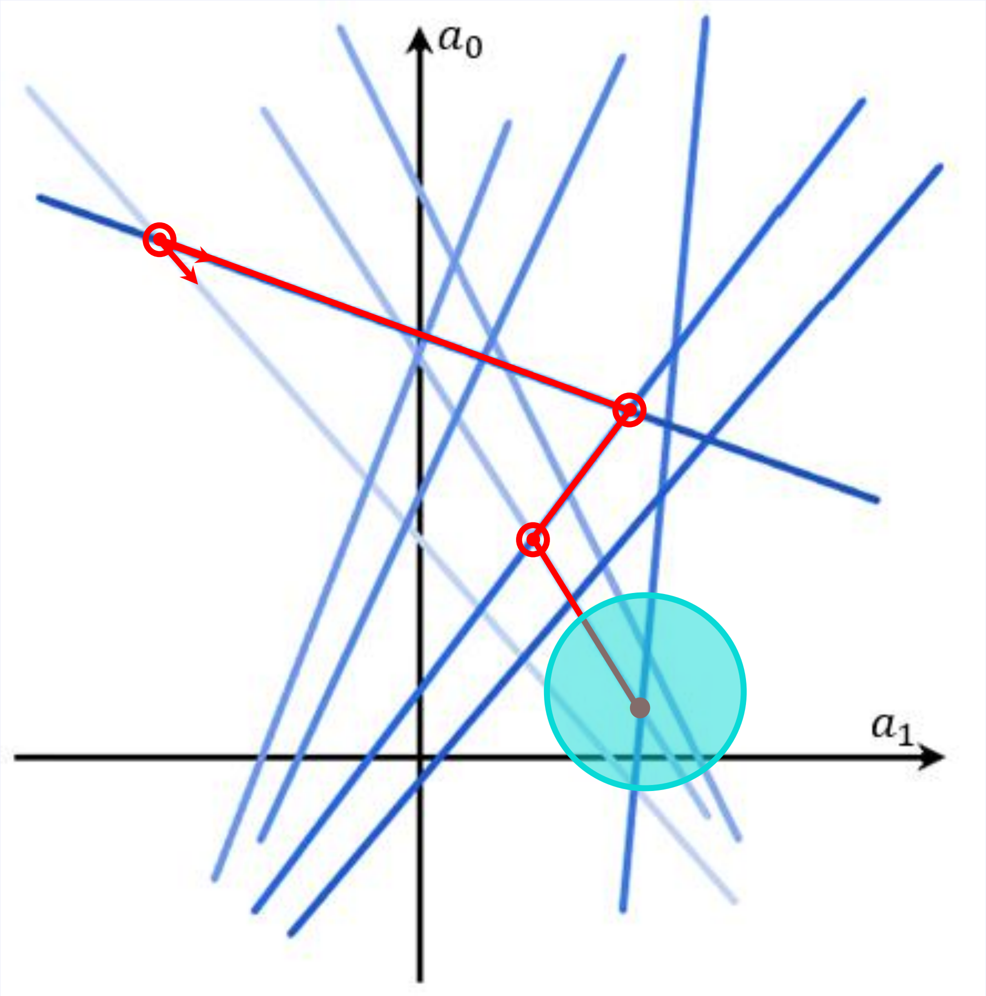
Вид функции Q(а) на узловой прямой; $T_{i_1}, T_{i_2}, \dots, T_{i_{n-m}}$ являются узловыми точками

Учет угла наклона



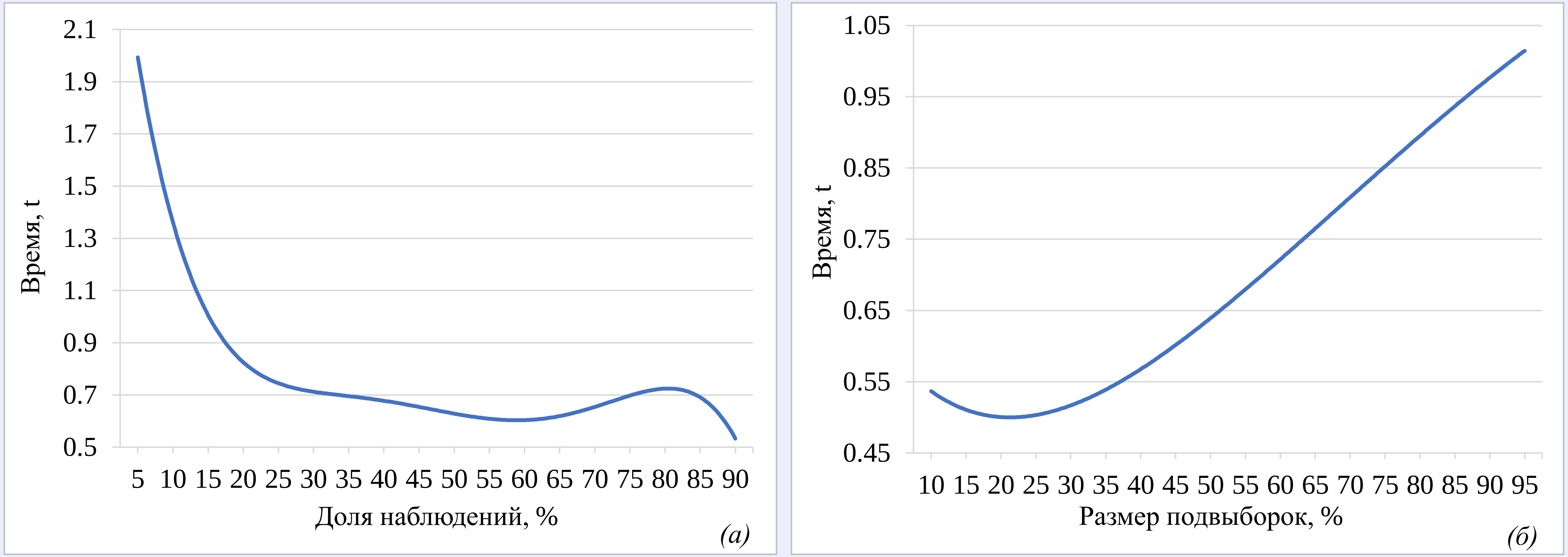
Вид функции Q(a) на узловой прямой; $T_{i_1}, T_{i_2}, \ldots, T_{i_{n-m}}$ являются узловыми точками

Учет «пучка»



Спуск по узловым прямым

Учет «пучка»



Сглаженные графики времени работы алгоритма при $m = 3, n = 250$: а – динамическое добавление доли выборки при $\alpha = 10$, б – первое приближение

Время вычисления

Покоординатный спуск		Градиентный спуск		Мод. градиентный спуск
$m^{2.1}n^{1.9} \cdot 0.00005$	$\longrightarrow$	$m^{2.6}n^{1.5} \cdot 0.0001$	$\longrightarrow$	$m^{2.9}n^{1.3} \cdot 0.00013$

$$\frac{0.5n^{0.4}}{m^{0.5}} \qquad \frac{0.8n^{0.2}}{m^{0.3}}$$

$$\boldsymbol{\frac{0.4n^{0.6}}{m^{0.8}}}$$

Симплекс-метод[6]

Прямая задача

$$\begin{cases} \mathbf{b}^T \tilde{\mathbf{y}} \to \min, \\ \mathbf{A}\tilde{\mathbf{y}} \geq \mathbf{C}, \\ \tilde{\mathbf{y}} \geq 0, \end{cases}$$

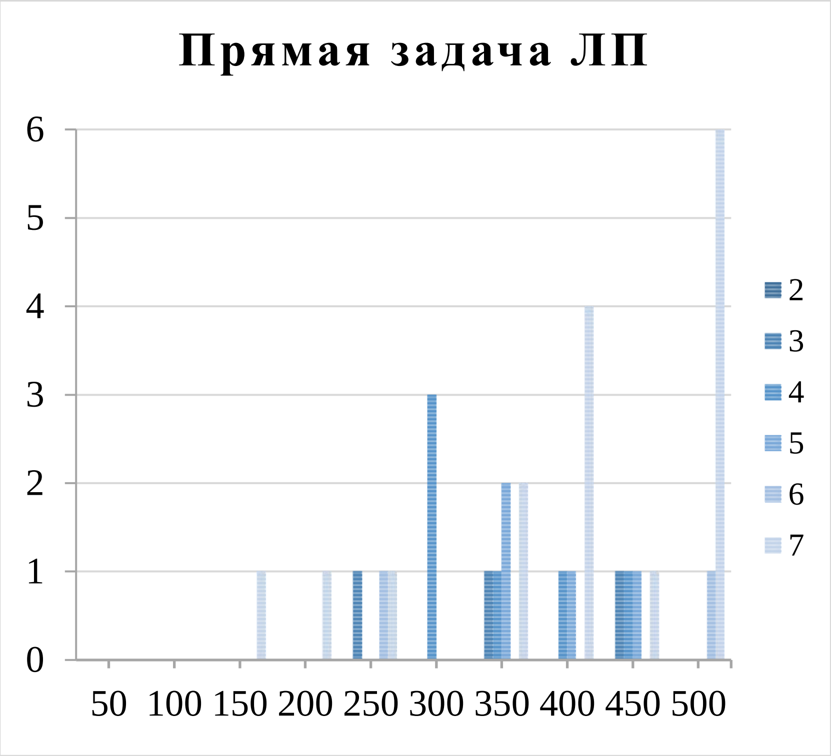
Двойственная задача

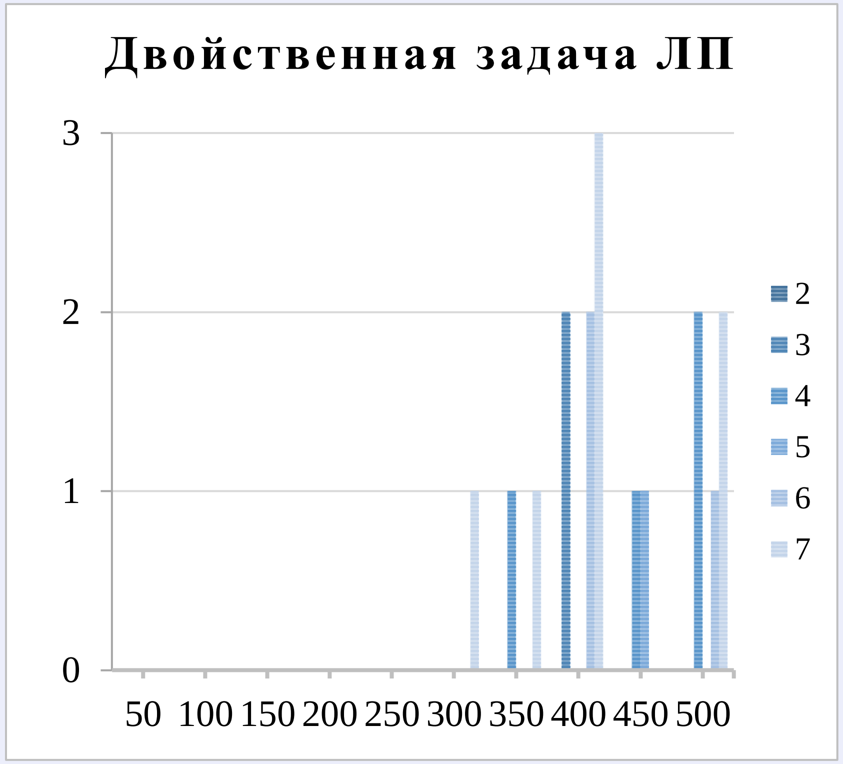
$$\begin{cases} \mathbf{C}^T \tilde{\mathbf{x}} \to \max, \\ A^T \tilde{\mathbf{x}} \leq \mathbf{b}, \\ \tilde{\mathbf{x}} \geq 0. \end{cases}$$

Базис	b_1	b_2	…	b_n	b_{n+1}	b_{n+2}	…	b_{n+2m-1}	b_{n+2m}	C
b_{n+2m+1}	z_1	0	…	0	$a_1^{(1)} x_{11}$	$-a_1^{(2)} x_{11}$	…	$a_m^{(1)} x_{1m}$	$-a_m^{(2)} x_{1m}$	y_1
b_{n+2m+2}	0	z_2	…	0	$a_1^{(1)} x_{21}$	$-a_1^{(2)} x_{21}$	…	$a_m^{(1)} x_{2m}$	$-a_m^{(2)} x_{2m}$	y_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$b_{2n+2m-1}$	0	0	…	z_n	$a_1^{(1)} x_{n1}$	$-a_1^{(2)} x_{n1}$	…	$a_m^{(1)} x_{nm}$	$-a_m^{(2)} x_{nm}$	y_n
b_{2n+2m}	z_1	0	…	0	$-a_1^{(1)} x_{11}$	$a_1^{(2)} x_{11}$	…	$-a_m^{(1)} x_{1m}$	$a_m^{(2)} x_{1m}$	$-y_1$
$b_{2n+2m+1}$	0	z_2	…	0	$-a_1^{(1)} x_{21}$	$a_1^{(2)} x_{21}$	…	$-a_m^{(1)} x_{2m}$	$a_m^{(2)} x_{2m}$	$-y_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
b_{3n+2m}	0	0	…	z_n	$-a_1^{(1)} x_{n1}$	$-a_1^{(2)} x_{n1}$	…	$-a_m^{(1)} x_{nm}$	$a_m^{(2)} x_{nm}$	$-y_n$
$F(X)$	c_1	c_2	…	c_n	c_{n+1}	c_{n+2}	…	c_{n+2m-1}	c_{n+2m}	F_0

[4] Голованов О.А., Тырсин А.Н. Спуск по узловым прямым и симплекс-алгоритм – два варианта регрессионного анализа на основе метода наименьших модулей // Заводская лаборатория. Диагностика материалов. 2023.

Симплекс-метод





Число отклонений решения прямой и двойственной задачи ЛП при помощи симплекс-метода от решения градиентным спуском по узловым прямым при $n=50,100,\dots,500$ и $m=2,3,\dots,7$

Метод проектирования градиента[7,8,9]

Для k-го шага матрица проектирования равна

$$P^{(k)} = E - X^{(k)\mathrm{T}}\left[X^{(k)}X^{(k)\mathrm{T}}\right]^{-1}X^{(k)},$$

Проекция градиента целевой функции

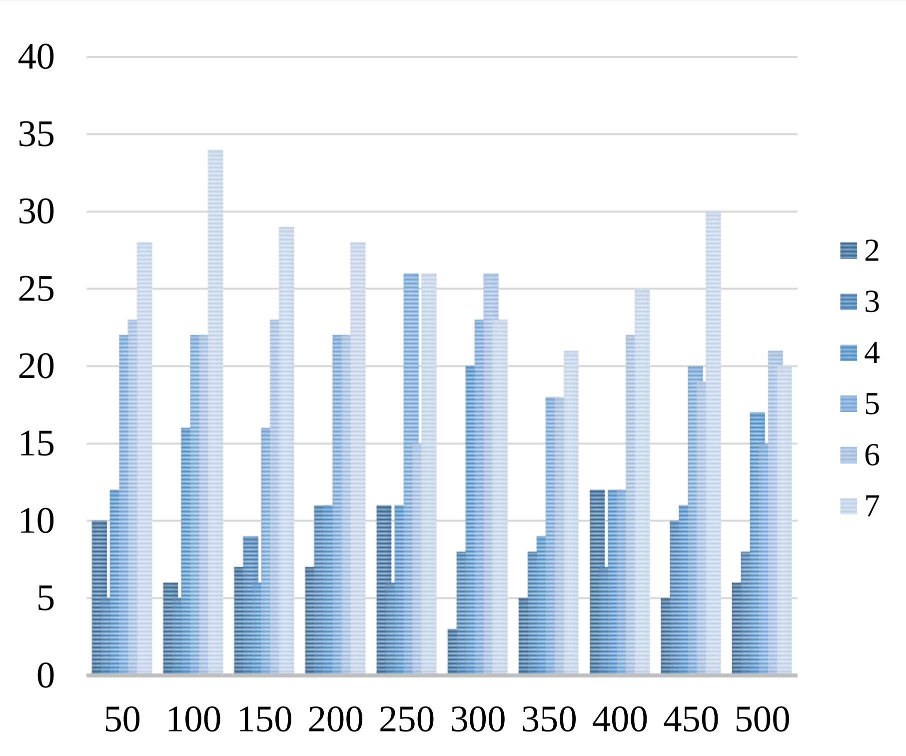
$$g^{(k)} = P^{(k)}y^{(k)}$$

Допустимая длина шага в направлении $g^{(k)}$

$$\alpha_* = \arg\max\left\{\alpha: -p_i \le \omega_i^{(k)} + \alpha g_i^{(k)} \le p_i, i \in S^{(k)}\right\}$$

Следующая точка

$$\omega_i^{(k+1)} = \omega_i^{(k)} + \alpha_* g_i^{(k)}, i \in S^{(k)}$$



Число отклонений метода проектирования градиента от решения градиентным спуском по узловым прямым при n=50,100,…,500 и m=2,3,…,7

[7] Панюков, А. В. Параметрическая идентификация квазилинейного разностного уравнения / А. В. Панюков, Я. А. Мезал // Вестник Южно-Уральского государственного университета. Серия: Математика. Механика. Физика. – 2019. – Т. 11, № 4. – С. 32-38. – DOI 10.14529/mmph190404.

[8] Мину, М. Математическое программирование: теория и алгоритмы / М. Мину. — М.: Наука. Гл. ред. физ.-мат. лит., 1990. — 488 с.

[9] Rosen, J.B. The gradient projection method for nonlinear programming, part 1: linear constraints / J.B. Rosen // Journal S.I.A.M. — 1960. — Vol. 8. pp. 181-217.

Сравнительный анализ времени вычисления

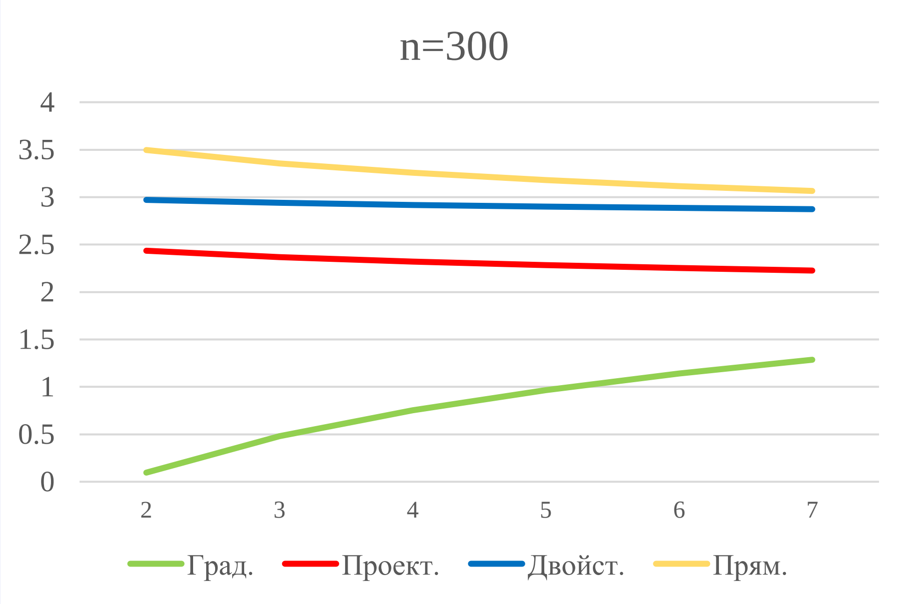
Мод. градиентный спуск
$$m^{2.9}n^{1.3} \cdot 0.00013$$

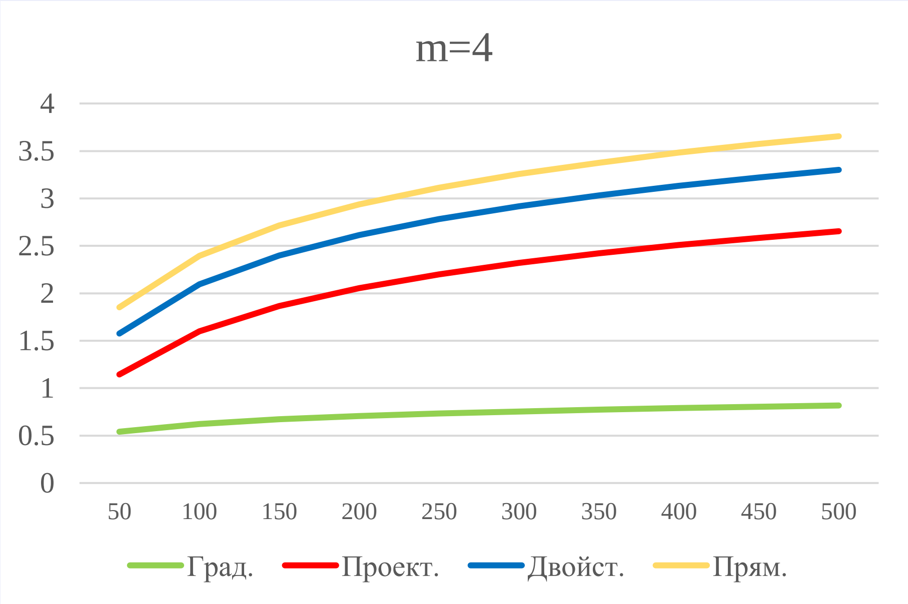
Прямой симплекс-метод
$$m^{0.001}n^{2.8} \cdot 0.0004$$

Двойственный симплекс-метод
$$m^{0.6}n^{2.7} \cdot 0.0001$$

Метод проект. градиента
$$m^{0.4}n^{2.5} \cdot 0.0001$$

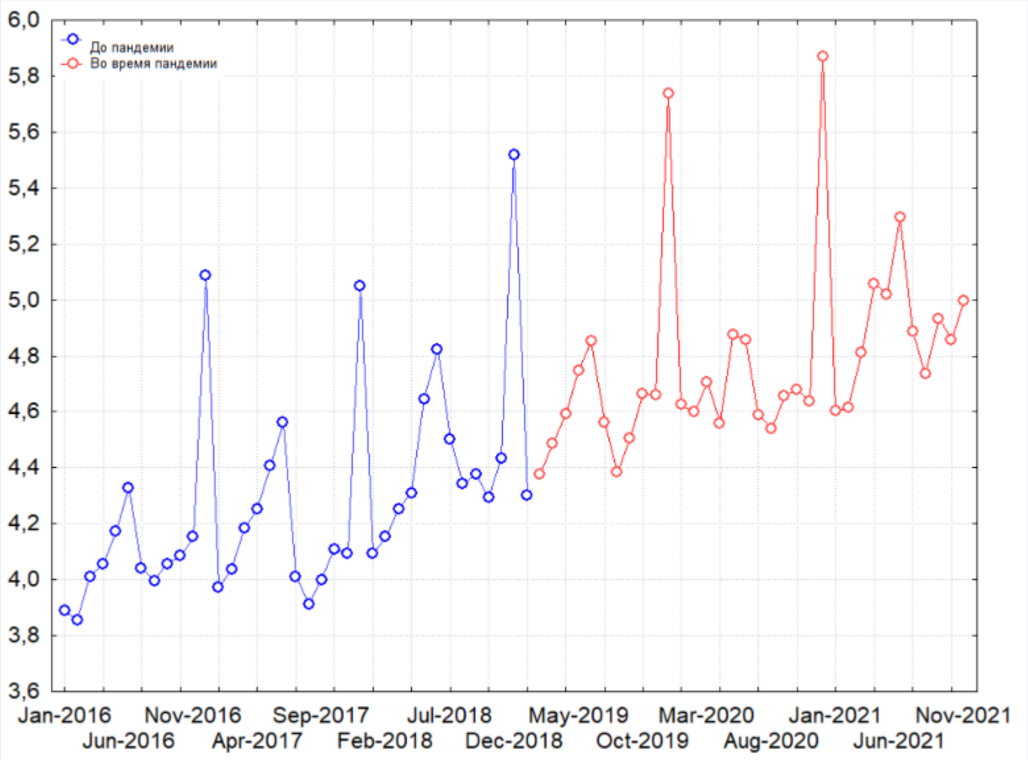
Метод наим. квадратов
$$m^{0.8}n \cdot 0.002$$





Десятичный логарифм от проигрыша методу наименьших квадратов при n=300 и m=4

Авторегрессия



$$\mathbf{Y}_t = \mathbf{a_0} + \mathbf{a_1}\mathbf{Y}_{\mathbf{t-1}} + \boldsymbol{\varepsilon}_t$$

где $\mathbf{Y}_t$ – фактические значения показателя, $\mathbf{Y}_{\mathbf{t-1}}$ – значение показателя с лагом в 1 период, **ε** – вектор случайных отклонений, **а** – вектор искомых коэффициентов

Особенности:

- Несимметричные выбросы (вероятность отклонения влево и вправо от нуля не равно ½);
- Зависимость случайных отклонений друг от друга и от объясняющей переменной.

Обобщенный метод наименьших модулей

ОМНМ-оценки для модели (1) находят как решение задачи[10]:

$$W(\mathbf{a}) = \sum_{i=1}^{n} \rho_{\text{ОМНМ}}(|y_i - \langle \mathbf{x}_\text{i}, \mathbf{a} \rangle|) \to \min_{\mathbf{a} \epsilon \mathbf{R}^m},$$

В отличие от целевой функции метода наименьших модулей, $W(\mathbf{a})$ не является явной выпуклой функцией, поэтому не всегда представляется возможным осуществить спуск к точному решению задачи, но МНМ и ОМНМ объединяет общее свойство – в обоих случаях решение находится в узловых точках.

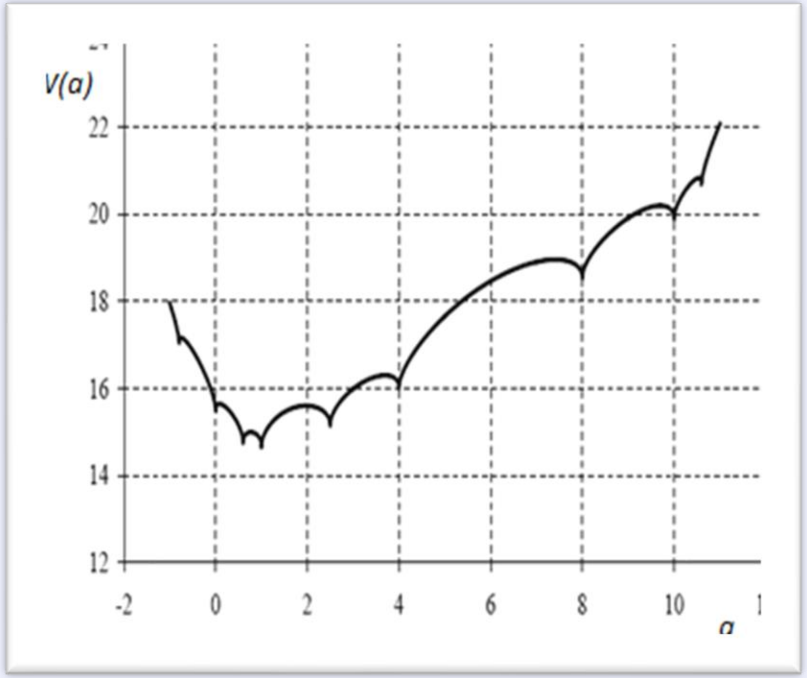
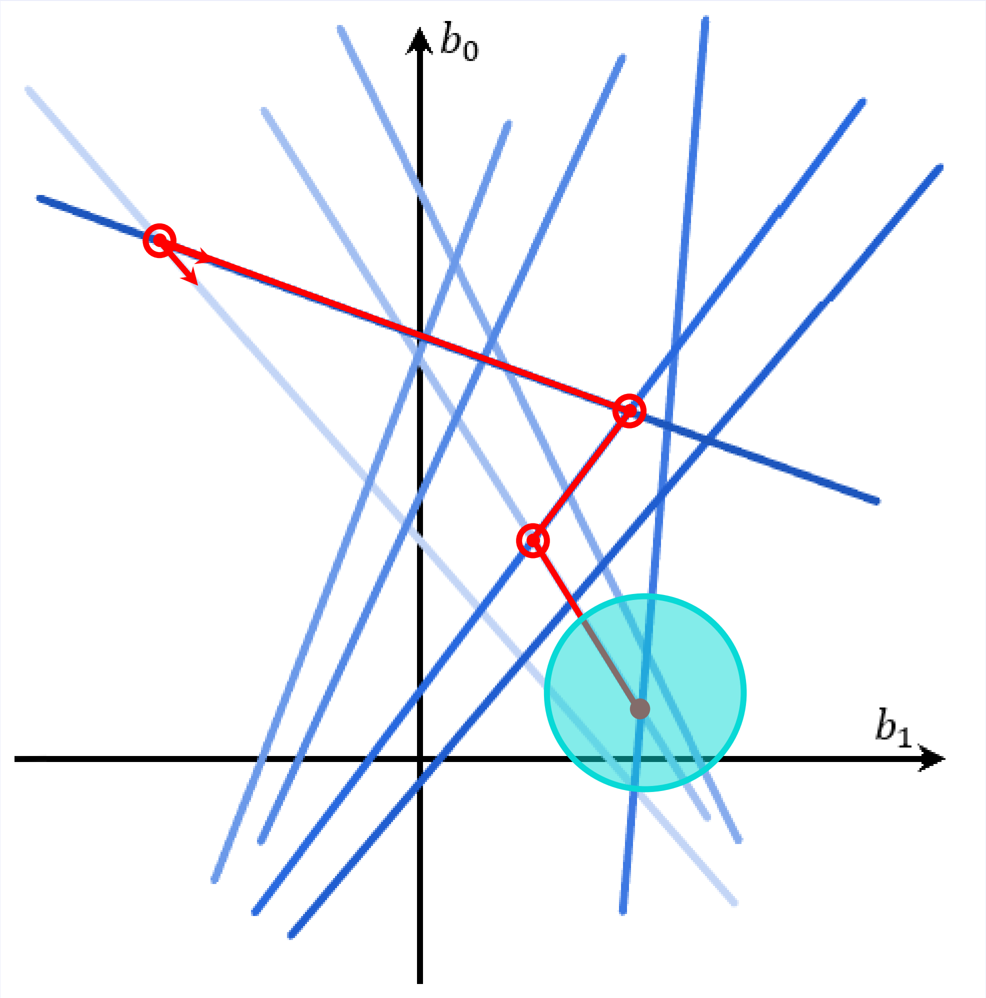


График целевой функции *W*(a) для *m*=1

[10] Тырсин, А.Н. Оценивание линейной регрессии на основе обобщенного метода наименьших модулей / А.Н. Тырсин, Л.А. Соколов // Вестник Самарского государственного технического университета. Серия Физико-математические науки. 2010. № 5(21). С. 134–142.



Спуск по узловым прямым

Оптимальная область решений

$p = 0.99$

n\m	2	3	4	5
50	84%	86%	87%	87%
100	47%	48%	48%	49%
150	33%	34%	34%	35%
200	26%	27%	27%	27%
250	22%	22%	22%	23%
300	19%	19%	19%	19%
350	16%	17%	17%	17%
400	15%	15%	15%	15%
450	13%	13%	14%	14%
500	12%	12%	12%	13%

$$\alpha_{99} = \frac{21.9}{n^{0.84}} m^{0.04}$$

$p = 0.95$

n\m	2	3	4	5
50	27%	35%	41%	46%
100	14%	18%	21%	24%
150	10%	12%	15%	17%
200	8%	9%	11%	13%
250	6%	8%	9%	10%
300	5%	7%	8%	9%
350	4%	6%	7%	8%
400	4%	5%	6%	7%
450	4%	4%	5%	6%
500	3%	4%	5%	5%

$$\alpha_{95} = \frac{7.05}{n^{0.93}} m^{0.57}$$

Оптимальная область решений

$p = 0.99$

n\m	2	3	4	5
50	39	42	44	45
100	43	46	48	50
150	45	48	51	53
200	47	50	53	55
250	48	52	55	57
300	50	53	56	58
350	51	54	57	59
400	52	55	58	61
450	53	56	59	62
500	53	57	60	63

$$\alpha_{99} = \frac{21.9}{n^{0.84}} m^{0.04}$$

$p = 0.95$

n\m	2	3	4	5
50	15	18	21	23
100	15	18	21	23
150	15	18	21	23
200	15	18	21	23
250	15	18	21	23
300	15	18	21	23
350	15	18	21	23
400	15	18	21	24
450	15	18	21	24
500	15	18	21	24

$$\alpha_{95} = \frac{7.05}{n^{0.93}} m^{0.57}$$

```
Введите полное название файла для анализа: 1.txt

Пожалуйста, проверьте правильность введенных данных!
Количество коэффициентов X: 2
Число наблюдений: 500000
Ваша функция примет следующий вид:
a0*1+ a1*72.4191 = 147.721

Какой уровень доверительной вероятности использовать 95 или 99%? (Введите 95 или 99)
95
Введите число итераций: 1000
Введите число наблюдений для одной итерации: 500

Значение целевой функции: 290.989
Конечные коэффициенты: 3.00671 1.99992
Среднее время обработки одной итерации: 0.093036
```

```
1.txt
Файл Правка Формат Вид Справка
1 72.4191 147.721
1 100.4 203.14
1 105.755 216.42
1 75.8414 154.723
1 79.3267 160.924
```

Консоль вызова программы

Конец доклада

Спасибо за внимание