

На связь роста и веса человека влияет много факторов: на то она и корреляция, а не функция. Один из важных факторов – это возраст человека. Немногим людям удастся сохранить свой вес таким, каким он был в молодости<sup>1</sup>. На вес человека также большое влияние оказывает его генетика (наследственность) и, конечно, образ жизни. У человека, как и у автомобиля (см. ниже) возраст определить несложно. А вот "пробег" человека, конкретно говоря, число шагов, которые он сделал за свою жизнь, тоже хорошо было бы измерять для более точного определения взаимозависимости веса и роста. Пробег замеряет спидометр автомобиля. Некоторые люди на прогулках или пробежках используют шагомер. Если читатель перейдет по ссылке, отмеченный в списке литературы [69], то он может узнать, как в среде Mathcad проводился более сложный регрессионный анализ – как определялась формула, по которой можно оценить стоимость подержанного автомобиля в зависимости от его возраста и пробега.

Автору тут вспоминается одна история, свидетелем которой он был в Западной Германии в 80-х годах прошлого века. Один немецкий профессор (автор был на стажировке в одном немецком университете) захотел быстро продать свой автомобиль за 2000 марок и выставил его по этой низкой цене на местном автомобильном "блошином" рынке. К нему довольно долго подходили потенциальные покупатели, осматривали машину, задавали вопросы, но все эти контакты оказывались безрезультатными. Но тут к нему подходит один довольно бойкий человек, покупает, как говорится, не глядя машину, меняет на ней... ценник и тут же на глазах изумленного профессора перепродает ее за 4000 марок... Этот перекупщик умел сходу правильно оценить стоимость подержанного автомобиля и понимал, что заниженная цена также плоха для торга, как и завышенная. Проецируя этот тезис на наших студентов, можно сказать, что слишком низкий рост также плох для жизни, как и слишком высокий...

Итак, автомобиль с его возрастом и пробегом!

Когда-то давно посетитель форум пакета Mathcad поместил на нем таблицу 14.1 и попросил форумчан подсказать, как из таблицы можно получить расчетную формулу для оценки стоимости подержанного автомобиля по его пробегу и возрасту.

Таблица 14.1

Данные о стоимости подержанного автомобиля определенной марки

---

<sup>1</sup> "Когда хорошее настроение, что-то удалось, нет посторонних и помещение подходящее, хочется раскинуть руки, загудеть и побежать по кругу, изображая самолет на вираже... Давно я не совершал такой "полет", а тут "совершил". Жена посмотрела и сказала, что раньше, в молодости я был легкомоторным самолетом, а теперь стал... широкофюзеляжным." ([www.anekdot.ru](http://www.anekdot.ru))

Возраст (лет)	Пробег (миль)	Цена (UD\$)	Возраст (лет)	Пробег (миль)	Цена (UD\$)	Возраст (лет)	Пробег (миль)	Цена (UD\$)
11,5	88000	1195	13,5	120000	495	7,5	41000	3495
13,5	103000	750	11,5	101000	895	13,5	124000	850
10,5	82000	1295	3,5	39000	4995	10,5	77000	1595
10,5	65000	1495	10,5	78000	1295	6,5	56000	3495
12,5	97000	800	6,5	52000	2695	12,5	83000	895
10,5	70000	1495	9,5	84000	1995	9,5	67000	2495
8,5	51000	2295	4,5	39000	3995	4,5	38000	3990
10,5	80000	1495	4,5	46000	3675	6,5	43000	3400
9,5	79000	1995	12,5	92000	795	13,5	92000	795
6,5	57000	2695	11,5	108000	975	11,5	78000	1295

Попытаемся сделать такой регрессионный анализ.

Таблица 14.1 перенесена в документ Mathcad Prime в виде трех векторов с размерностями времени, длины и стоимости, объединенных в таблицу – см. рис. 14.22.

Область

*US\$ :=* □

<i>Возраст</i> <i>(yr)</i>	<i>Пробег</i> <i>(mi)</i>	<i>Цена</i> <i>(US\$)</i>
11.5	88000	1195
13.5	103000	750
10.5	82000	1295
10.5	65000	1495
12.5	97000	800
10.5	70000	1495
8.5	51000	2295

Рис. 14.22. Таблица Mathcad Prime с данными о подержанных автомобилях

Статистический анализ данных принято начинать с построения различного рода графических зависимостей, которые могут подсказать возможные пути обработки данных. На рисунках 14.23-14.25 можно видеть графики, построенные по точкам из "автомобильной" таблицы 14.1.

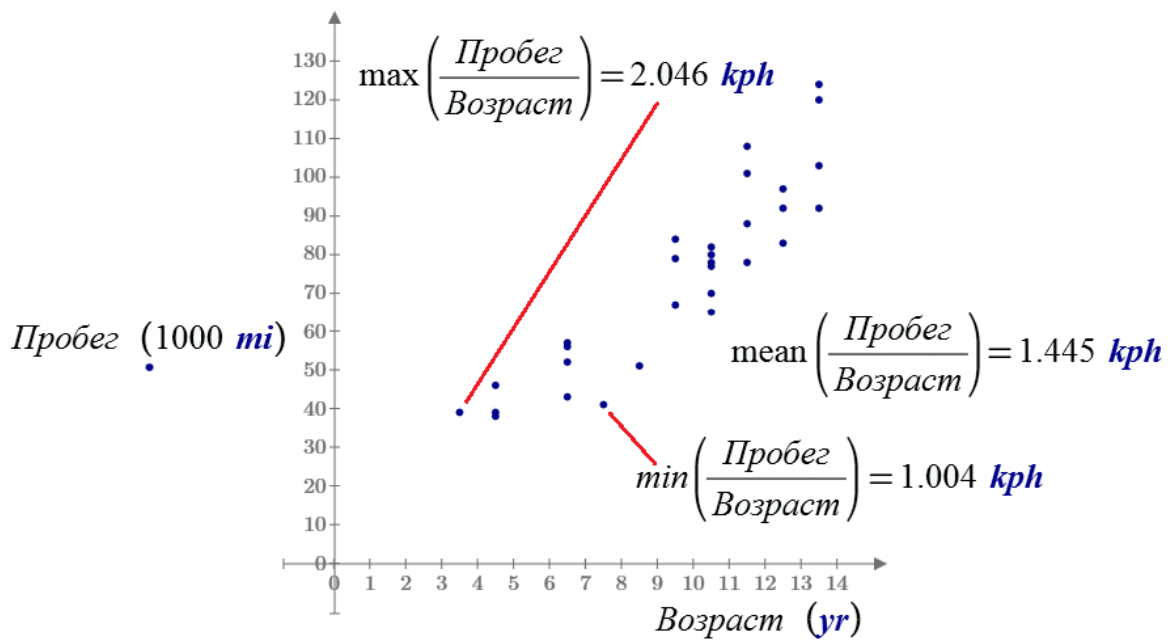


Рис. 14.23. Корреляция пробега автомобиля и его возраста

Если пробег автомобиля разделить на его возраст, то мы получим среднюю скорость автомобиля за время его эксплуатации. К этой дроби можно применить три встроенные функции  $\max$ ,  $\text{mean}$  и  $\min$ , возвращающие минимальное, среднее и максимальное значения вектора. Из рисунка 14.23 видно, что корреляция пробега автомобиля и его возраста сосредоточена у прямой линии с наклоном 1.445 км/ч, проходящей через начало координат. В нашей выборке минимальная и максимальная скорости такие: 1.004 км/ч (7.5 лет и 41 000 миль пробега) и 2.046 км/ч (3.5 года и 39 000 миль). Если провести две прямые линии через начало координат и эти две точки, то получится некий "эксплуатационный" клин, внутри которого лежат наши 30 точек выборки. Отсюда вывод – если вам предлагают купить подержанный автомобиль, параметры которого не попадают в этот "клин" (1-2 км/ч), то нужно подумать, прежде чем приобретать такую машину, которую либо слишком интенсивно эксплуатировали (скорость больше 2 км/ч), либо она непонятно почему простаивала (скорость меньше 1 км/час), либо у нее... подкручен спидометр. Возраст машины скрыть (подделать) намного трудней.

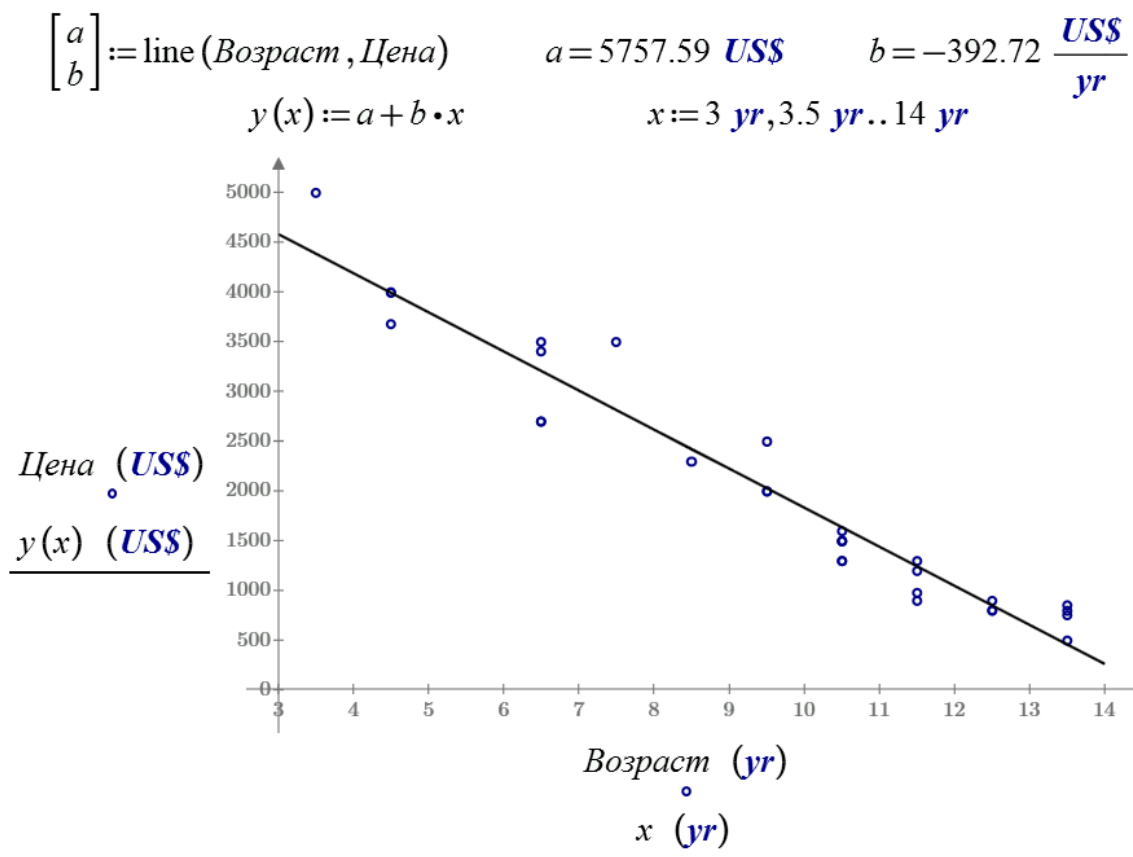
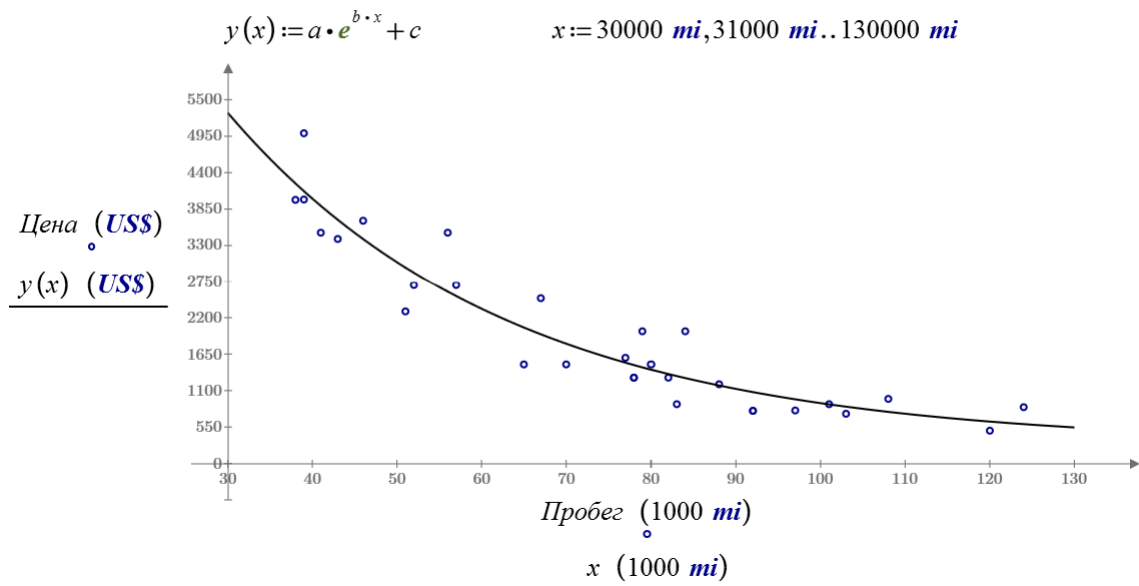


Рис. 14.24. Корреляция цены автомобиля и его возраста

Цена автомобиля от возраста зависит линейно  $a + b \cdot x$  – см. 14.24. Никакой другой более сложной зависимости из точек, разбросанных на рисунке 14.24, выудить невозможно. В нашем случае после каждого года пробега с цены машины нужно будет скидывать в среднем по 393 доллара. Цена же нового автомобиля составляет примерно 5750 долларов.

Зависимость цены автомобиля от пробега (рис. 14.25) более сложная – она меняется экспоненциально:  $a \cdot e^{b \cdot x} + c$ . При такой зависимости ( $b < 0$ ) цена машины с пробегом никогда не упадет до нуля, чего не скажешь о зависимости цены от возраста (рис. 14.24), когда в районе 14-15 лет автомобиль становится "бесценен" в одном из двух смыслов: либо у него теперь нулевая цена (наш смысл) либо автомобиль стал... "бесценным" музейным экспонатом.

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} := \text{expfit} \left( \text{Пробег}, \text{Цена}, \begin{bmatrix} 10000 \text{ } \mathit{US\$} \\ -0.001 \frac{1}{\mathit{mi}} \\ 100 \text{ } \mathit{US\$} \end{bmatrix} \right) \quad a = 12291.48 \text{ } \mathit{US\$} \quad b = -2.991 \cdot 10^{-5} \frac{1}{\mathit{mi}} \quad c = 293.091 \text{ } \mathit{US\$}$$



**Рис. 14.25.** Корреляция цены автомобиля и его пробега

Попробуем теперь создать нашу искомую функцию двух аргументов – возраста и пробега, по которой можно будет оценить стоимость подержанного автомобиля. Ее вид могут подсказать две отдельные функции одного аргумента (см. рис. 14.24 и 14.25). Искомая функция будет иметь четыре параметра (a, b, c и d – см. первую строку в расчете на рис. 14.26) и представлять собой комбинацию линейной и экспоненциальной зависимостей. Необходимо только рассчитать значения параметров a, b, c и d, при которых точки, зафиксированные в таблице 14.1 и на графиках 14.23-14-25, оказались бы вблизи уже не линии, а поверхности. Для этого создается вспомогательная функция, возвращающая сумму квадратов отклонений точек от поверхности – см. второй оператор на рис. 14.26. Мы такую операцию уже проводили при регрессионном анализе роста и веса студентов – см. рис. 14.7.

$$z(\text{Возраст}, \text{Пробег}, a, b, c, d) := a + b \cdot \text{Возраст} + c \cdot e^{d \cdot \text{Пробег}}$$

$$\Sigma\Delta 2(a, b, c, d) := \sum_{i=1}^{\text{last}(\text{Цена})} \left( \text{Цена}_i - z(\text{Возраст}_i, \text{Пробег}_i, a, b, c, d) \right)^2$$

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} := \begin{bmatrix} 4000 \text{ US\$} \\ -300 \frac{\text{US\$}}{\text{yr}} \\ 2000 \text{ US\$} \\ 0 \cdot \frac{1}{\text{mi}} \end{bmatrix} \quad \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} := \text{minimize}(\Sigma\Delta 2, a, b, c, d)$$

$$a = 4000 \text{ US\$} \quad b = -300 \frac{\text{US\$}}{\text{yr}} \quad c = 2000 \text{ US\$} \quad d = 0 \frac{1}{\text{mi}}$$

Рис. 14.26. Неудачная попытка работы с функцией minimize

На рисунке 14.26 сделана попытка расчета параметров  $a$ ,  $b$ ,  $c$  и  $d$  через поиск минимума суммы квадратов отклонений точек от поверхности. Она оказалась неудачной – функция `minimize` вернула значения первого приближения. Дело в том, что в пакет Mathcad Prime был встроен новый алгоритм работы функции `minimize`. А новое нередко бывает недоработанным. Пришлось обратиться к функции `minerr`, которая уже давно встроена в Mathcad и в которой выявлены возможные ошибки и недоработки – см. рис. 14.27. Мы эту функцию использовали в этюде 4 и показывали, как ею можно заменить функцию `minimize` – см. рис. 4.17 и 4.20.

Начальные приближения
Ограничения
Решатель

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} := \begin{bmatrix} (4.454 \cdot 10^3) \text{ } \alpha \\ -(9.091 \cdot 10^{-6}) \frac{\alpha}{s} \\ (1.255 \cdot 10^4) \text{ } \alpha \\ -(3.95 \cdot 10^{-8}) \frac{1}{m} \end{bmatrix}$$

$$\Sigma \Delta 2(a, b, c, d) = 0 \text{ } \text{US\$}^2$$

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} := \mathbf{minerr}(a, b, c, d) = \begin{bmatrix} (4.454 \cdot 10^3) \text{ } \alpha \\ -9.092 \cdot 10^{-6} \frac{\alpha}{s} \\ (1.256 \cdot 10^4) \text{ } \alpha \\ -3.95 \cdot 10^{-8} \frac{1}{m} \end{bmatrix}$$

$$a = 4453.81 \text{ } \text{US\$} \quad b = -286.91 \frac{\text{US\$}}{\text{yr}}$$

$$c = 12555.641 \text{ } \text{US\$} \quad d = -6.358 \cdot 10^{-5} \frac{1}{\text{mi}}$$

Рис. 14.27. Удачная попытка работы с функцией `minerr`

Но и с функцией `minerr` пришлось повозиться: давать первое предположение, получать ответ, видеть, что это еще не минимум и переносить (копировать) ответ на место первого предположения. Контроль же за качеством минимизации велся через построения четырех декартовых графиков с горизонтальными маркерами, на которых один из четырех параметров



задается переменной диапазона, а три остальные фиксируются как константы – см. рис. 14.28-14.31: в точке минимума функции, как известно, значения всех ее частных производных должны быть равны нулю.

$$\Delta := 1\% \quad av := a \cdot (1 - \Delta), a \cdot (1 - \Delta) + \frac{a \cdot (1 - \Delta)}{3000} .. a \cdot (1 + \Delta)$$

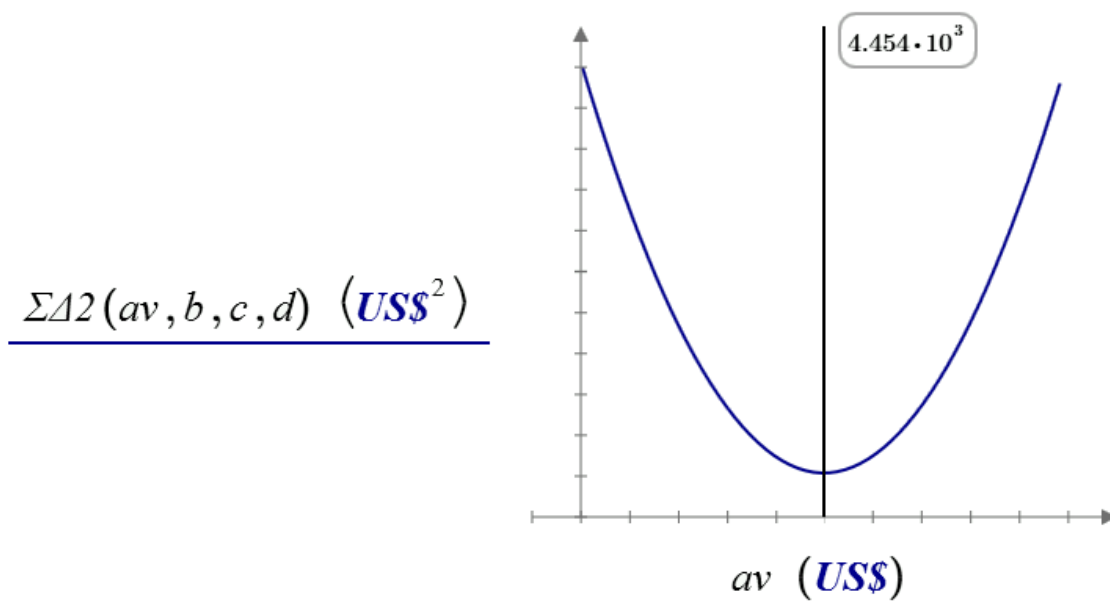


Рис. 14.28. Проверка точки минимума по параметру а

$$bv := b \cdot (1 - \Delta), b \cdot (1 - \Delta) + \frac{b \cdot (1 - \Delta)}{3000} .. b \cdot (1 + \Delta)$$

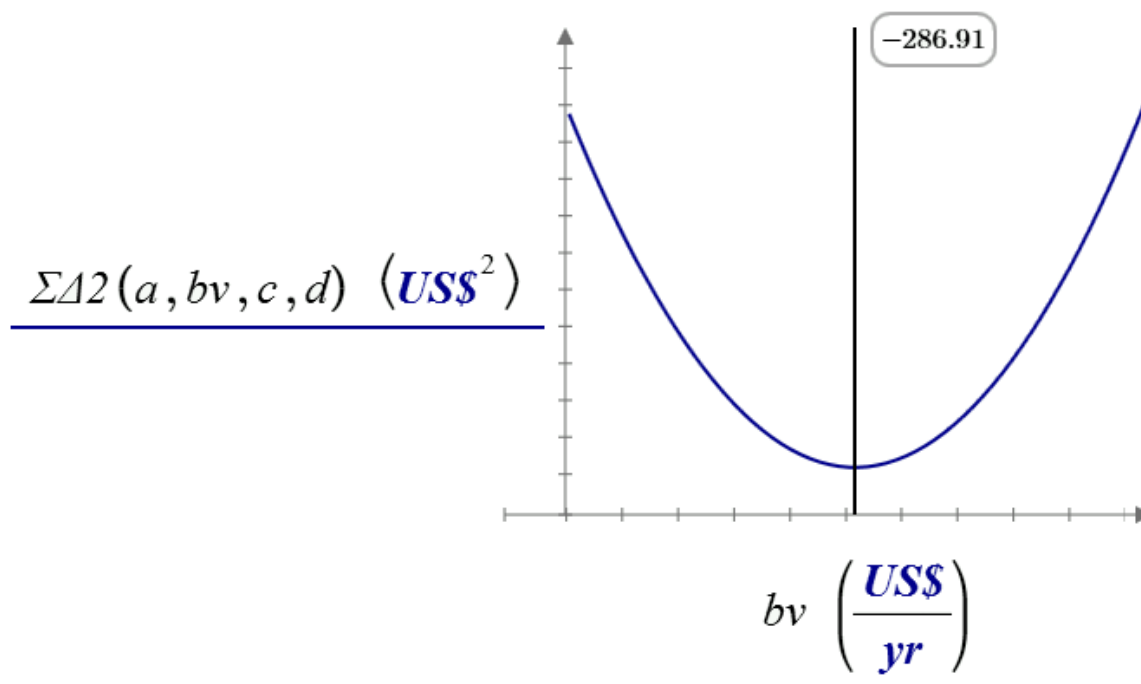


Рис. 14.29. Проверка точки минимума по параметру б

$$cv := c \cdot (1 - \Delta), c \cdot (1 - \Delta) + \frac{c \cdot (1 - \Delta)}{3000} .. c \cdot (1 + \Delta)$$

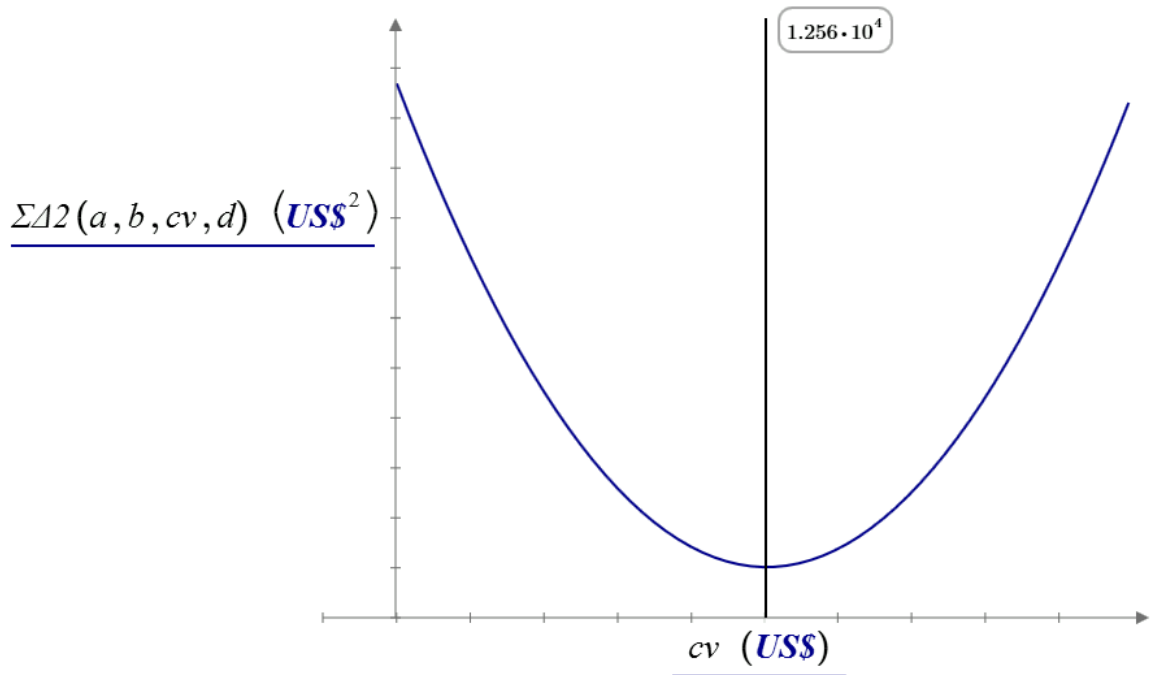


Рис. 14.30. Проверка точки минимума по параметру  $c$

$$dv := d \cdot (1 - \Delta), d \cdot (1 - \Delta) + \frac{d \cdot (1 - \Delta)}{3000} .. d \cdot (1 + \Delta)$$

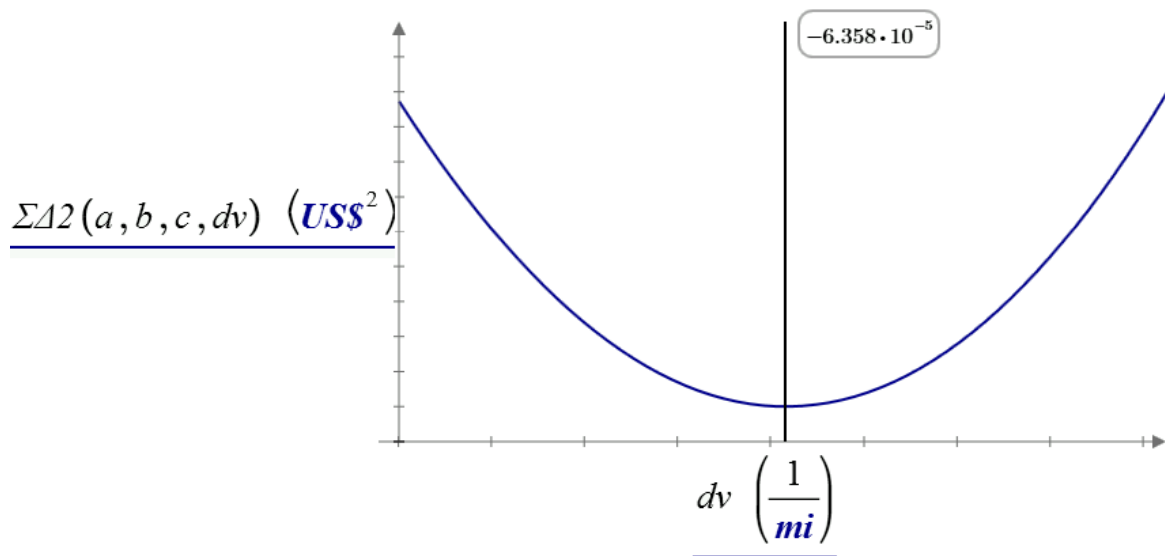


Рис. 14.31. Проверка точки минимума по параметру  $d$

Честно говоря, было желание изъять из книги рисунок 14.26 и не рассказывать об этой недоработке Mathcad Prime. Тем более, следует ожидать, что к моменту выхода книги в свет, появится новая версия этого пакета, где эта ошибка будет исправлена. Но рисунок был оставлен в книге и вот почему.

Нередки ситуации, когда пользователь программы видит, что в ней что-то не так и по умолчанию считает себя виновником ошибки, несмотря на то, что все было сделано вроде бы правильно. А ошибка сидит в самой программе. При таких сомнениях следует обратиться на форум Mathcad и узнать мнения других людей. Так было и сделано, и данная ошибка в функции `minimize` была зафиксирована и принята разработчиками. Кроме того, нужно помнить, что многие инструменты Mathcad и других математических программ продублированы. Это позволяет решать задачу разными способами, "уворачиваясь" от возможных недоработок.

Так или иначе, задача об оптимальных значениях параметров  $a$ ,  $b$ ,  $c$  и  $d$  регрессионного выражения для цены подержанного автомобиля была решена и найденную формулу можно использовать при оценке машин с пробегом – см. рис. 14.32, где вводятся значения переменных `МойВозраст` и `МойПробег`, рассчитывается цена автомобиля в долларах и в процентах от цены нового с нулевым пробегом и возрастом<sup>2</sup>, а также показана данная точка среди точек статистической выборки.

---

<sup>2</sup> Тут нужно еще учесть тот факт, что сразу после продажи реальная цена автомобиля падает на 20-30%.

$МойВозраст := 9 \text{ yr}$      $МойПробег := 110000 \text{ km}$

$z(МойВозраст, МойПробег, a, b, c, d) = 2034 \text{ US\$}$

$$\frac{z(МойВозраст, МойПробег, a, b, c, d)}{z(0 \text{ yr}, 0 \text{ mi}, a, b, c, d)} = 12\%$$

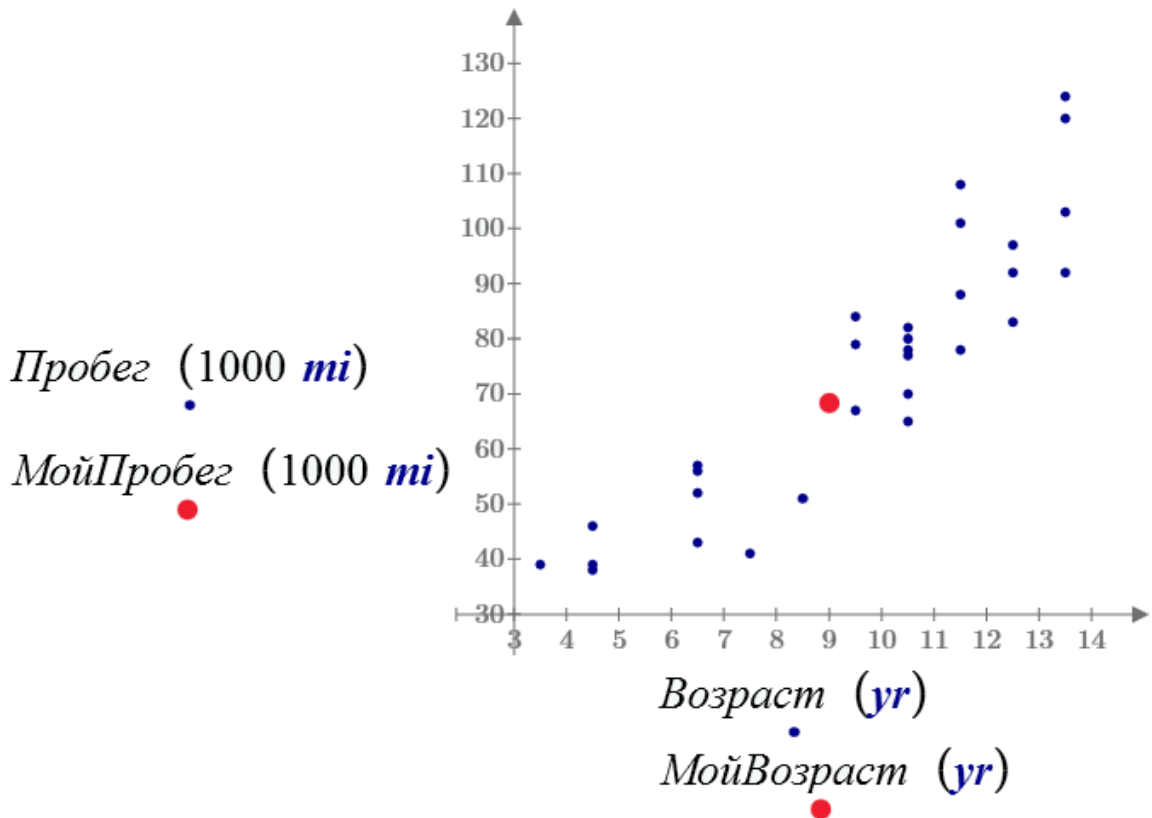


Рис. 14.32. Расчет цены подержанного автомобиля

Еще об одном сомнении. В расчете на рис 14.27 был найден минимум функции, а на рисунках 14.28-14-21 показана графическая проверка этого решения. Но не ясно, является ли этот минимум глобальным и нет ли других значения параметров  $a$ ,  $b$ ,  $c$  и  $d$ , при которых сумма средних квадратичных отклонений будет еще ближе к нулю.

Чтобы развеять сомнения, нужно применить иные методы поиска минимума функции – например, генетический метод, авторский сайт которого показан на рис. 14.33. На этом сайте есть ссылка на описание метода. Этот метод (эго реализация для Mathcad сделана В.Р.Сабаниным) позволяет найти глобальный минимум функции в заданной области ее аргументов. Так из рисунка видно, что функция пользователя двух аргументов имеет два локальных минимума и глобальный минимум, который и был найден.

<http://twtmas.mpei.ac.ru/mas/Worksheets/Minimum-3.mcd>

$f(x, y) := \frac{3*(x-1)*e^{-(x^2+(y+1)^2)}+10*(0.2*x-x^3-y^5)*e^{-(x^2+y^2)}+e^{-(x^2+y^2)^2}}{3}$

$x_b := -3$        $x_e := 3$       Окончательная точность  
 $y_b := -3$        $y_e := 3$        $\varepsilon := 0.00001$

Разброс графика       $\Delta := 5$       Начальное число популяций  
 $\mu := 50$

Search Global Minimum of  $f(x, y)$

$x_{opt} = -0.010239$        $y_{opt} = 1.5806761$        $f(x, y)_{min} = -8.1157996$

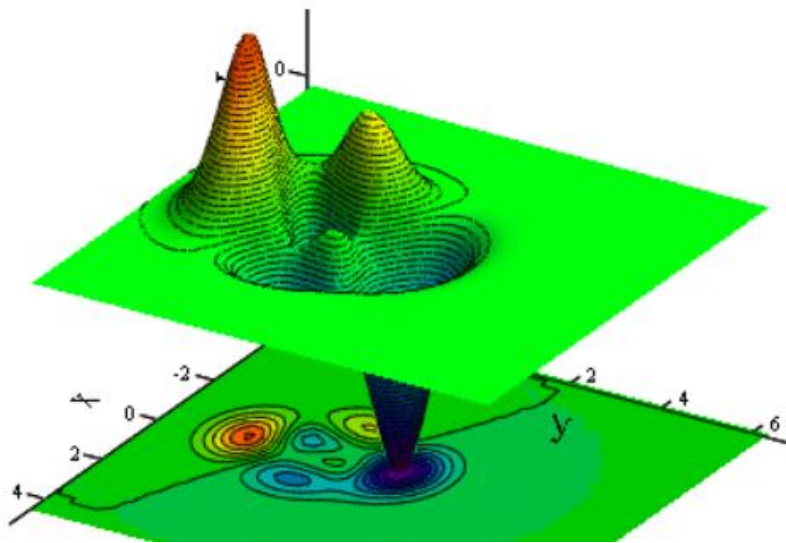


Рис. 14.33. Поиск глобального минимума по генетическому алгоритму

Статистическая обработка данных по цене подержанного автомобиля позволила создать Mathcad-сайт для оценки данной корреляции – см. рис. 14.34.

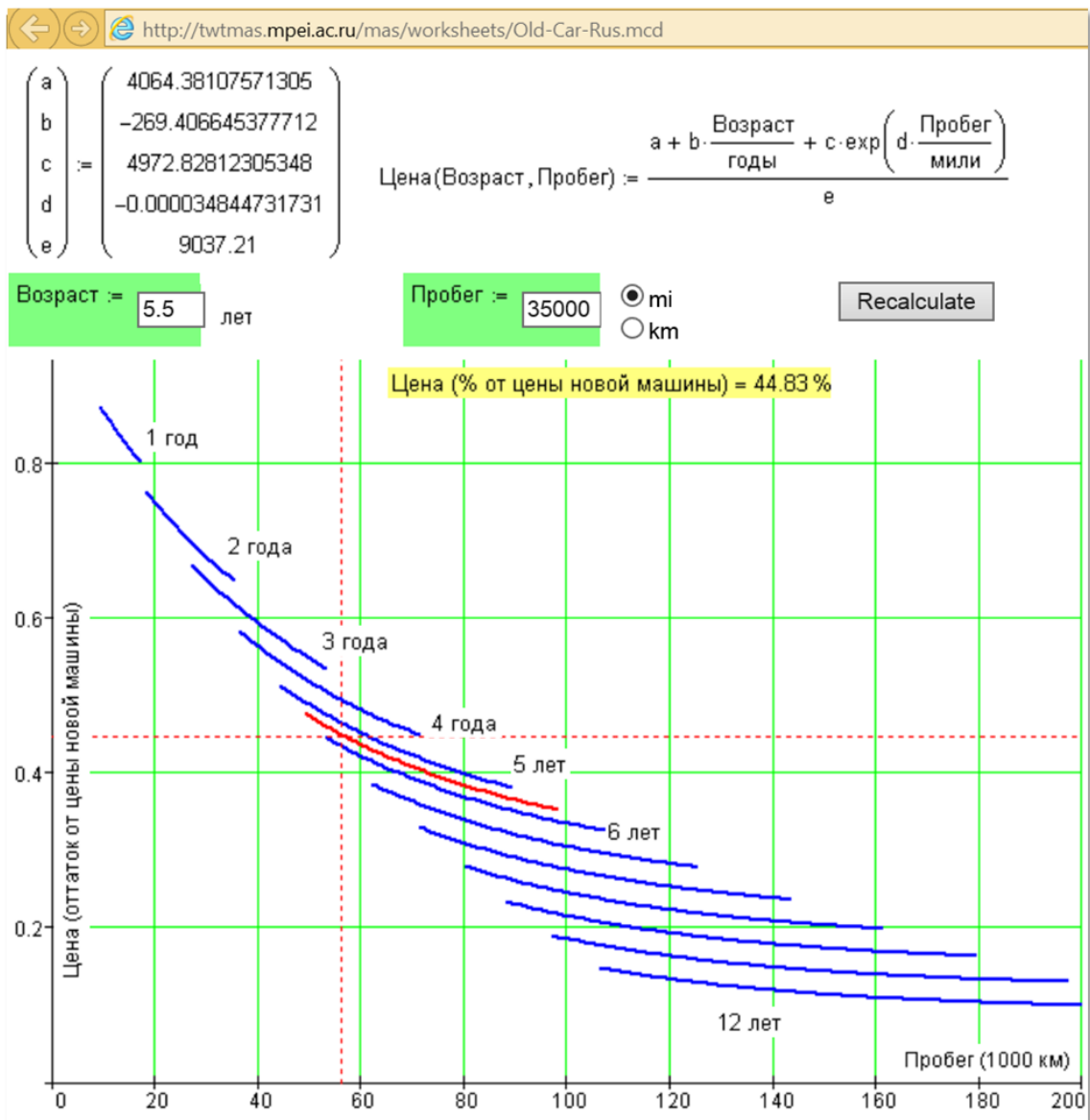


Рис. 14.34. Сайт по оценке стоимости подержанного автомобиля

По графику на рис. 14.34 можно не только получить оценочную стоимость старого автомобиля, но и отсеять заведомо сомнительные варианты, зафиксированные на рис. 14.23 (параметры вне скоростного клина), когда пробег автомобиля не соответствует его возрасту. Mathcad-документ, по которому велись расчеты, можно "скачать" по адресу: <ftp://twt.mpei.ac.ru/ochkov/Auto>.

Один из авторов книги два раза приобретал автомобиль по схеме treid-in, когда в стоимость новой машины входит стоимость старой, сдаваемой на комиссию. Так вот, оба раза цена старого автомобиля, предлагаемая салоном, отличалась не более, чем на 5% от той, которая была рассчитана с помощью сайта, показанного на рис. 14.34.