Занятие 6. Мистер Х Или Немного статистики

Из числа всей её челяди самым замечательным лицом был дворник Герасим, мужчина двенадцати вершков роста, сложенный богатырем и глухонемой от рожденья.

И.С. Тугенев «Муму»

Какой-то <...> огромный вершков двенадцати господин, тоже необычайно толстый, чрезвычайно мрачный и молчаливый и, очевидно, сильно надеявшийся на свои кулаки.

Ф.М. Достоевский «Идиот»

А фельдъегерь уж там, понимаете, и стоит: трехаршинный мужичина какой-нибудь, ручища у него, можете вообразить, самой натурой устроена для ямщиков, — словом, дантист эдакой...

Н.В. Гоголь «Мёртвые души»

Для начала несколько определений с такой оговоркой.

Нижеприведенные определения, честно говоря, можно и не давать. Современный читатель, если видит в книге что-то непонятное, может сразу самостоятельно сделать поиск в интернете и найти справку.

Итак, определения.

Функция (от лат. functio — исполнение, осуществление) — это закон f, по которому каждому элементу x из одного числового множества ставится в соответствие некоторый определенный элемент y из другого числового множества, так что $y = f(x)$.

Корреляция (от лат. correlatio — соотношение) — вероятностная или статистическая зависимость, не имеющая строго функционального характера (см. выше) из-за невозможности точно учесть влияние множества одновременно меняющихся факторов.

Регрессия (от лат. regressio — обратное движение) — такая зависимость среднего значения какой-либо величины от некоторой другой величины или от нескольких величин, что, в отличие от функциональной зависимости $y = f(x)$, одному и тому же значению переменной x могут соответствовать в зависимости от случая различные значения величины y.

Тренд — основная тенденция изменения чего-либо.

Гистограмма — способ представления табличных данных в графическом виде — в виде столбчатой диаграммы.

Интерполяция (от лат. interpolation — изменение, переделка) — способ нахождения промежуточных значений величины по имеющемуся дискретному набору известных значений этой величины, т. е. восстановление (обычно приближенное) функции (см. выше) по её нескольким известным значениям.

А теперь к делу.

В сентябре каждого года в родном вузе автора – в МЭИ проводятся два конкурса: «Мисс Первокурсница» и «Мистер Первокурсник». Мероприятие, по мнению автора, довольно сомнительное и по форме, и по содержанию. Подобные конкурсы лучше проводить не в начале, а в конце учебы в вузе, определяя лучшего выпускника по интерьеру – по внутреннему содержанию, а не по экстерьеру – по внешним данным.

Отметим также, что в финалы подобных конкурсов обычно, но далеко не всегда попадают красавицы-раскрасавицы, из которых очень трудно выделить какую-то одну. Отсюда слезы, обиды и даже судебные тяжбы.

Но под экстерьером можно понимать не какую-то там абстрактную красоту, которая непонятно как должна спасти мир, а более объективные данные – вес и рост участников конкурса, например.

6.1. Статистическая выборка студентов

Когда-то перед лекцией на тему «Регрессионный анализ» в рамках курса «Информационные технологии» автор подбирал пример *статистической выборки* для такого анализа. Но когда он «взошел на кафедру» и взглянул на аудиторию (рис. 6.1), то он понял, что эта выборка находится перед его глазами [1].

Рис. 6.1. Студенты автора

Была проведена необычная перекличка студентов. Юноши через СМС или иным способом сообщали лектору свой вес и рост, а также номер своей учебной группы. Студентки по понятным причинам в таком опросе не участвовали.

Полученная информация, но не вся, а выборка из выборки – данные юношей одной учебной группы курса заносилась в среде математической программы SMath в матрицу *Data* с двумя строками и двадцатью столбцами (см. рис. 6.2), из которой затем изымались два вектора с именами *Вес* и *Рост*. Логичнее было бы, конечно, матрицу *Data* сделать не «горизонтальной», а «вертикальной» – с двумя столбцами и 20 строками. В этом случае её бы не приходилось транспонировать – задействовать оператор, внешне похожий на оператор степени Т. Но «горизонтальная» матрицы компактнее помещается в расчётном документе.

Также стоит отметить, что матрица *Data* по своей сути – это реляционная *база данных* с двумя полями и 20 записями.

В векторах получилось по 20 элементов (переменная n). Эти массивы чисел послужили хорошей «затравкой» для лекции.

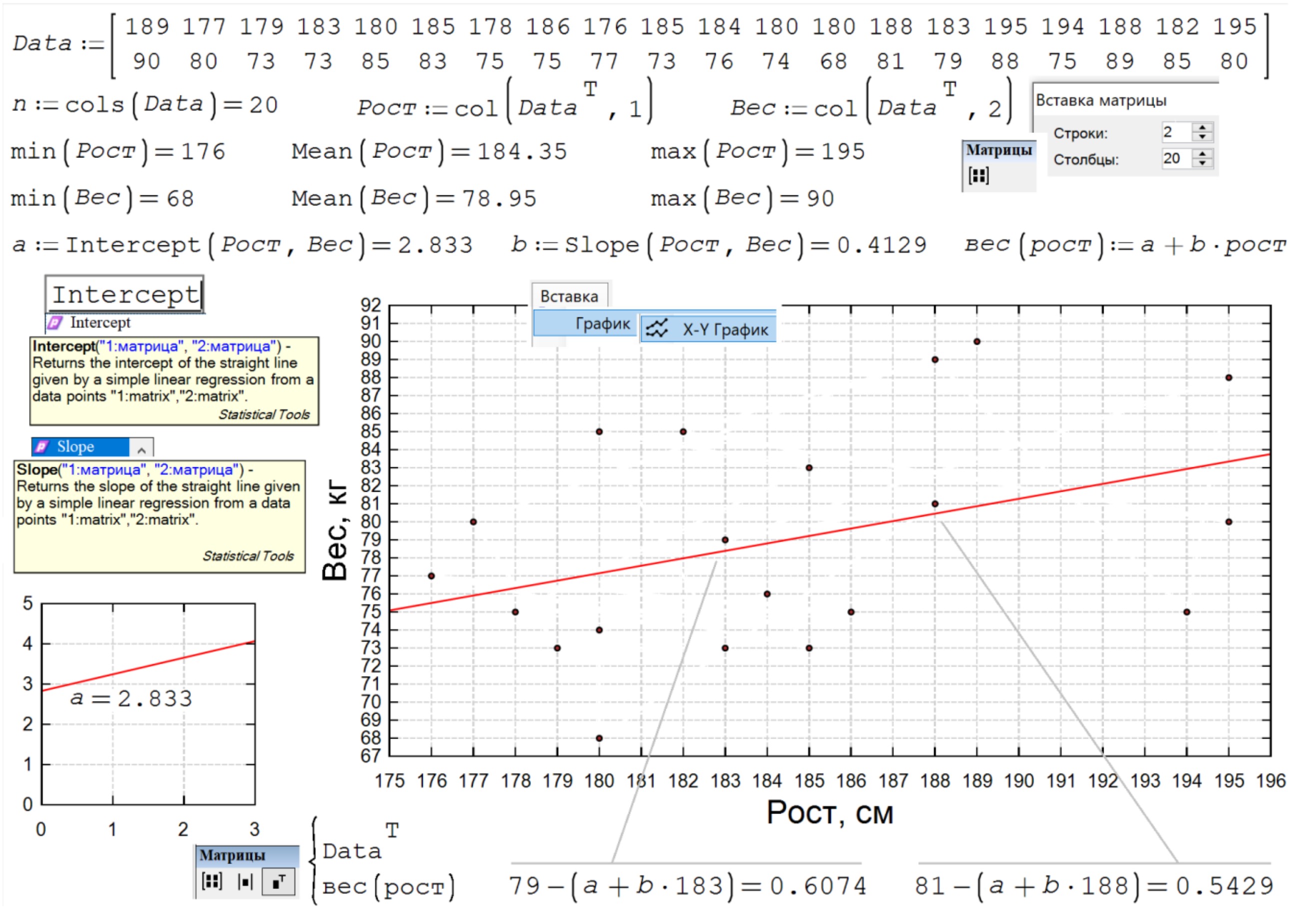


Рис. 6.2. Линейная регрессия

На точки графика на рис. 6.2 в докомпьютерную и докалькуляторную эру накладывали прозрачную линейку, двигали её вверх-вниз (изменение параметра a – см. ниже), поворачивали влево-вправо (изменение параметра b), пытаясь такими действиями «поймать» *линию тренда* – сделать ручной, в буквальном смысле этого слова, *регрессионный анализ*. В расчёте на рис. 6.2 эту работу делают две функции – `Intercept` и `Slope`, входящие в дополнение Statistical Tools к пакету SMath – см. рис. 6.3.

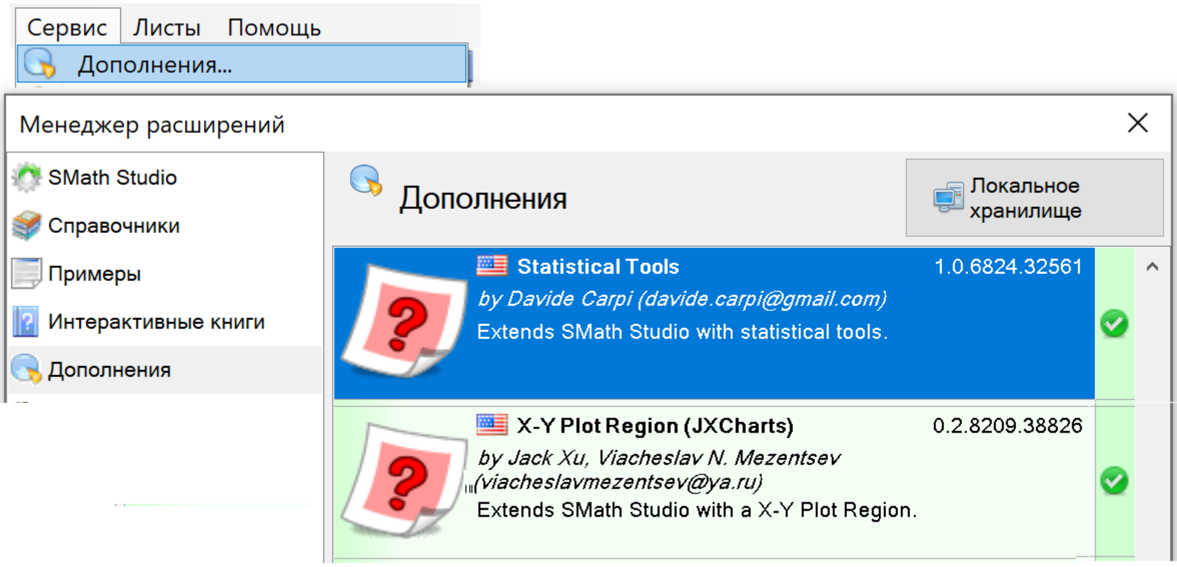


Рис. 6.3. Два дополнения к пакету SMath

Следует сразу отметить, что работа с прозрачной линейкой даст неверный результат, если некоторые точки на графике фактически являются двойными, тройными и т.д.: несколько человек в выборке имеют одинаковый рост и вес. В этом случае можно не дублировать такие одинаковые точки в матрице `Data`, а ввести третью строку у этой матрицы (с числами 1, 2, 3 и т.д.), хранящей вес данного элемента выборки. Но в этом случае расчёт, показанный на рис. 6.2, будет работать неправильно – см. первое задание читателям.

Если дополнение Statistical Tools не подгружено к пакету SMath, то необходимо перейти от локального хранилища к галерее online, найти его там и загрузить. Такой переход проводится щелканьем мышки по иконке «Локальное хранилище», открывающей короткий список с ещё одной позицией – Галерея онлайн (рис. 6.3).

Найденные параметры a и b, входящие в уравнение прямой линии тренда (а другую зависимость здесь увидеть довольно сложно), имеют четкий смысл. Эта линия пересекает ось ординат в точке с ординатой a (см. левый нижний угол на рис. 6.2 с некой бессмысленной *экстраполяцией*), а один сантиметр роста среднестатистического студента прибавляет ему b килограмм веса.

Несколько слов о графике на рис. 6.2. Такого типа плоской графики нет в ядре SMath – его тоже нужно подгрузить из интернета (см. рис. 6.3). Если график типа X-Y Plot Region вставить в расчёт и заполнить его поле ввода транспонированной матрицей `Data`, то мы получим «каляку-маляку», показанную вверху рисунка 6.4. Пакет SMath проставит на графике точки, координаты которых хранятся в отдельных строках матрицы `Data`, и соединит их отрезками прямых. Почему прямых? А потому, что по умолчанию использован метод Lines. Если метод заменить на Splines, то «каляка-маляка» станет гладкой – потеряет острые углы. Мы эту опцию использовали не в плане форматирования нашего графика, а для того, чтобы лишний раз показать, как SMath строит графики гладких функций – синусоиды, например. Пакет SMath разбивает интервал графика на 100 отрезков (см. позицию Points на рис. 6.5), табулирует точки на этом отрезке и соединяет их отрезками прямых. Если школьник или студент на занятиях по математике будет вручную таким манером строить график, то эмоциональный преподаватель выгонит его с занятий и будет топать ногами и улюлюкать ему вслед. Мы график той же синусоиды строим более интеллектуально – отмечаем на графике точки пересечения синусоиды с осью абсцисс (нули функции), находим точки минимума и максимума, а затем качественно, а не количественно рисуем кривую, помня о периоде синусоиды. Прямая строится нами вообще предельно просто – на графике проставляются две точки, которые соединяются прямой линией.

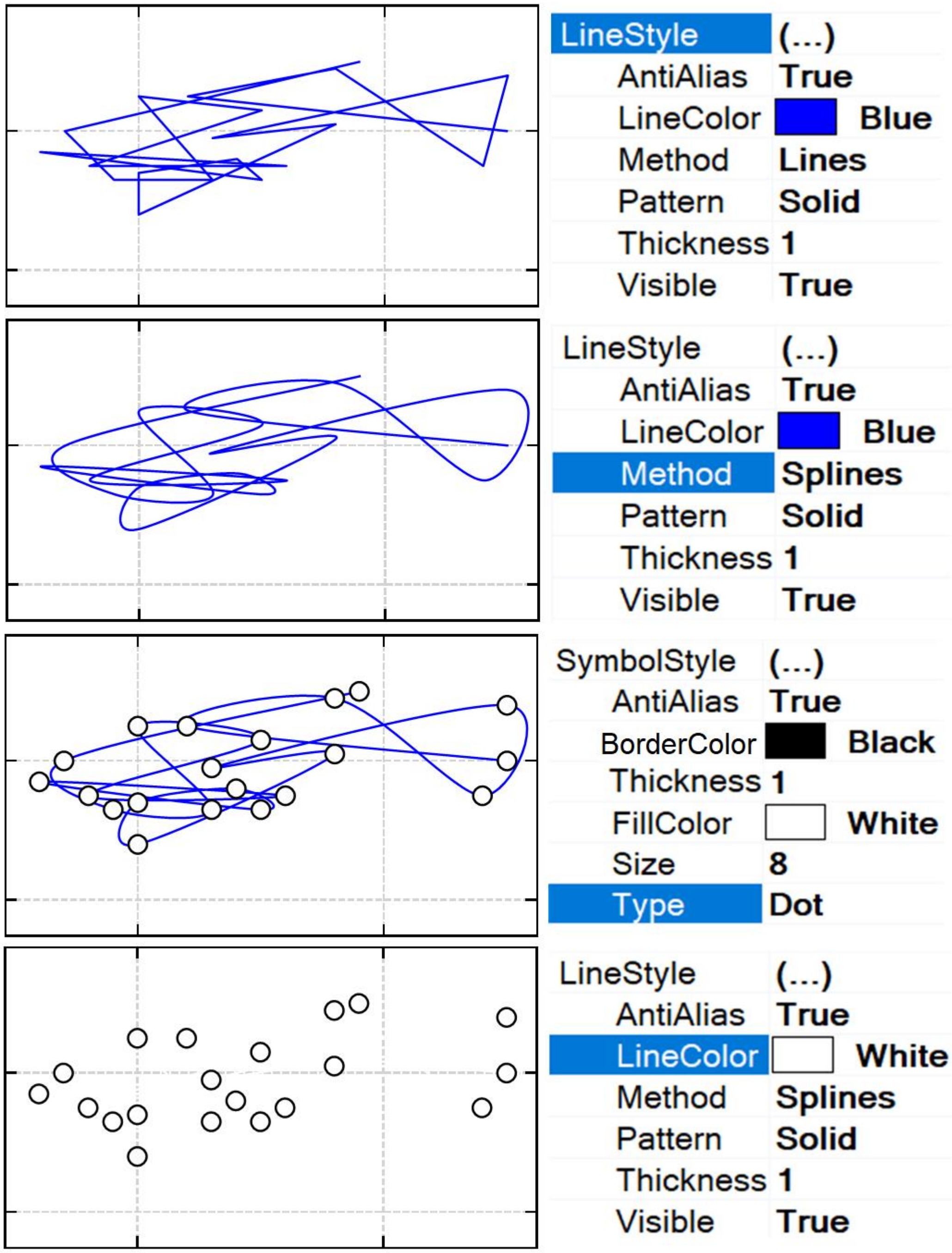


Рис. 6.4. Форматирование X-Y графика

Окна форматирования, показанные на рис. 6.4, вызываются двойным щелчком мыши по графику. После этого появится окно, показанное на рис. 6.5. Два нижних графика на рис. 6.4 с элементами форматирования показывают, как невидимые точки превращаются в кружочки (Dot) размером (Size) 8 условных единиц с черным кантом (BorderColor) и белым заполнением (FillColor). Замысловатая кривая, соединяющая точки, пропадет, если её сделать белой (LineColor) или прозрачной (Transparent, а не White). А это нам и нужно – см. точки на рис. 6.2.

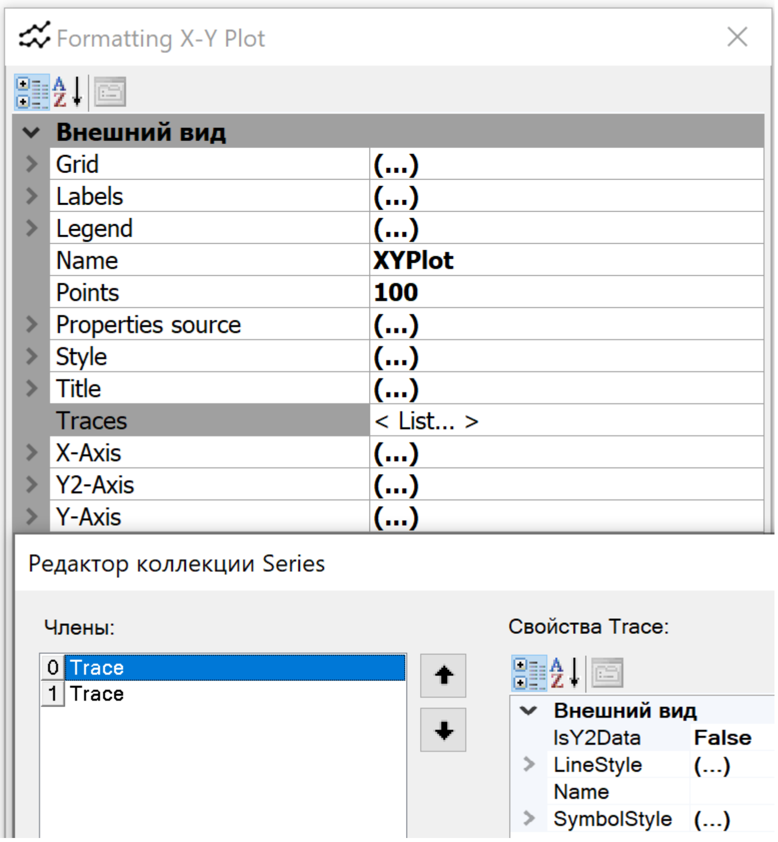


Рис. 6.5. Окно форматирования X-Y графика

Стоит обратить внимание на строку IsY2Data False (рис. 6.5). Замена слова False на слово True позволит ввести в график вторую ось ординат. В матрицу `Data` расчёта на рис. 6.2, можно ввести третью строку, хранящую, к примеру, размер обуви студента. Вторая ось ординат позволит провести ещё одну линию тренда – корреляцию роста человека и размера его ноги. В [2, 3] описано такое статистическое исследование по отношению к размеру головного убора.

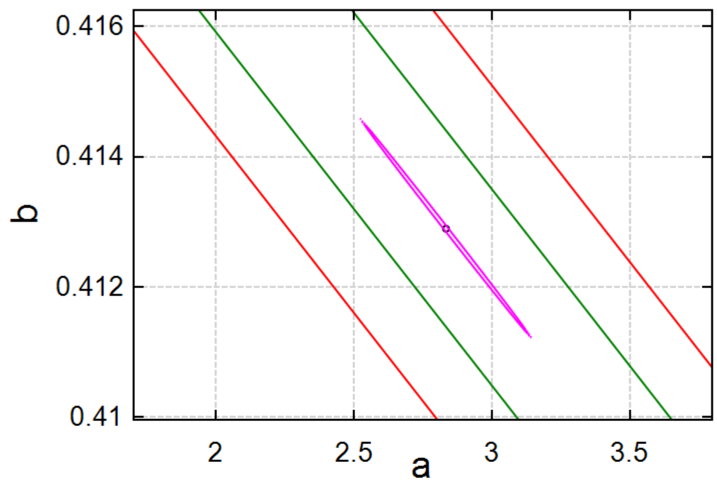
Но вернемся к нашему линейному регрессионному анализу – к конкурсу «Мистер X».

Две точки оказались самыми близкими к прямой тренда. Это и будут победители нашего конкурса. А кто из владельцев этих точек будет Мистером X, а кто Вице-мистером X? Для ответа на этот вопрос достаточно подсчитать *отклонение* этих точек от прямой – см.

последнюю строку на рис. 6.2. Мистер Х, то бишь Мистер Первокурсник – это юноша ростом в 188 см и весом в 81 килограмм, а вице-мистер Х – это юноша ростом в 183 см и весом в 79 килограмм!

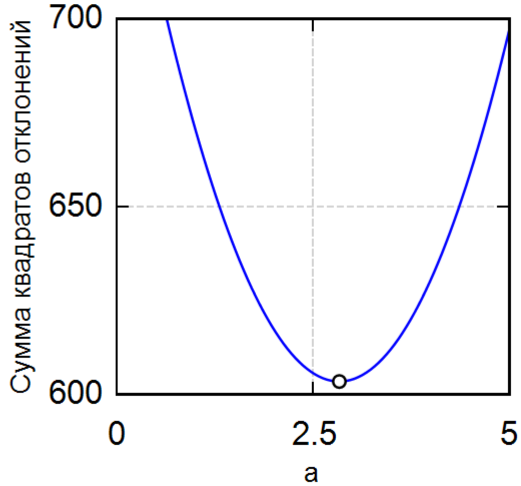
Если мы посчитаем такие отклонения для всех точек, возведем эти отклонения в квадрат и просуммируем эти квадраты, то мы получим число 603.5, показанное на первой строке расчёта на рис. 6.6. Если б все точки были бы на линии тренда, то эта сумма, как можно легко сообразить, была бы равна нулю. А так она равна какому-то положительному числу. При всяком другом расположении линии тренда, это число будет возрастать и это легко доказать. Следовательно мы имеем *минимум суммы квадратов отклонений*! Функции `Intercept` и `Slope` решают задачу линейной регрессии методом *наименьших квадратов*. В этом можно убедиться, взглянув на тот же рис. 6.6, где, во-первых, создана и просчитана функция пользователя с именем *СКО* (*сумма квадратов отклонений* точек от прямой), а во-вторых, построены графики (контурный график – графики сечений поверхности горизонтальными плоскостями), показывающие, что точка с координатами a = 2.833 и b = 0.4129– это точка минимума. Эту точку можно рассчитать и по-другому – найти корень системы двух уравнений: равенство нулю частных производных функции *СКО*, что показано на рис. 6.6 ниже контурного графика. Ещё ниже показаны два сечения поверхности функции с именем *СКО*.

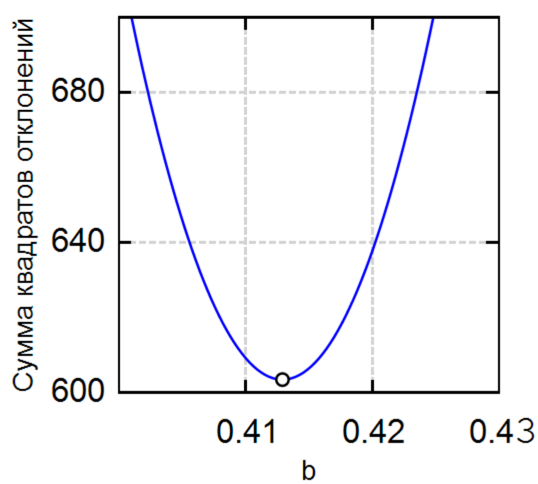
$$\text{СКО}(a, b) := \sum_{i=1}^{n} \left(\text{Вес}_{i} - \left(a + b \cdot \text{Рост}_{i}\right)\right)^{2} = 603.5$$



[a b "." 3]
СКО(x, y) − 610
СКО(x, y) − 605
СКО(x, y) − 603.455

$$\text{roots}\left(\begin{bmatrix} \frac{d}{d\,a}\text{СКО}(a, b) = 0 \\ \frac{d}{d\,b}\text{СКО}(a, b) = 0 \end{bmatrix}, \begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} 2.833 \\ 0.4129 \end{bmatrix}$$





СКО(x, b)
[a СКО(a, b) "." 5]

СКО(a, x)
[b СКО(a, b) "." 5]

Рис. 6.6. Графический анализ функции *СКО*

Отклонения возводятся во вторую степень для того, чтобы правильно учесть точки, находящиеся под графиком. Здесь можно было бы обойтись и абсолютным значением. Но эта функция в отличие от возведения во вторую степень имеет на графике острый угол, что затрудняет дальнейшую аналитическую работу с ней – взятие, например, производной.

Нюансы данного регрессионного анализа рассматривались на форуме пользователей SMath по адресу: https://en.smath.com/forum/yaf_postsm82699_Why-I-have-errors.aspx

А вот как ещё можно поступить с выборкой студентов. Девушек (они в белой форме) и юношей (черная форма) построили в несколько рядов по росту – см. рис. 6.7. Получилась некая живая *гистограмма*, отображающая распределение людей по росту с интервалом в один дюйм. Это отмечено на табличках, расположенных впереди рядов: первая цифра – футы, а вторая – дюймы.

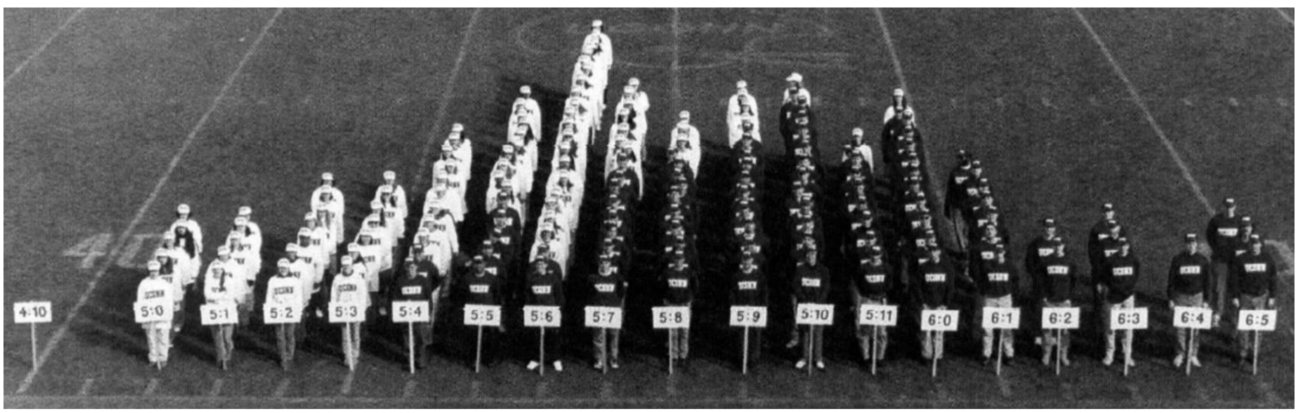


Рис. 6.7. Живая гистограмма

Если отойти от американского интервала (дюйм) и перейти к европейскому интервалу (сантиметр), то задний край нашей живой гистограммы окажется ещё более неровным. А там по идее должен быть гладкий статистический колокол – кривая какого-то (нормального?) распределения.

Здесь самое время обратиться нам к эпиграфам этого занятия.

Рост в старой России измеряли вершками – см. главу 16 в [4]. В аршине 16 вершков, а в сажени три аршина. Длина сажени в разное время была разной, пока её в середине девятнадцатого века не привязали к семи английским футам – см. рис. 6.8, где введены три пользовательских единицы длины.

$$сажень := 7\ ft = 213.36\ cm$$ Семь футов под килем!

$$аршин := \frac{сажень}{3} = 71.12\ cm$$

$$вершок := \frac{аршин}{16} = 4.445\ cm$$

$$Рост\_Герасима := 12\ вершок = 53.34\ cm$$???

$$2\ аршин = 142.2\ cm \qquad 3\ аршин = 213.4\ cm$$

$$Рост\_Герасима := 12\ вершок + 2\ аршин = 195.6\ cm$$!

Рис. 6.8. Старые русские меры длины

Но два аршина роста взрослого человека не указывались по умолчания (см. два первых эпиграфа) потому, что рост нормального взрослого мужчины находился между двумя и тремя аршинами – см. рис. 6.9.

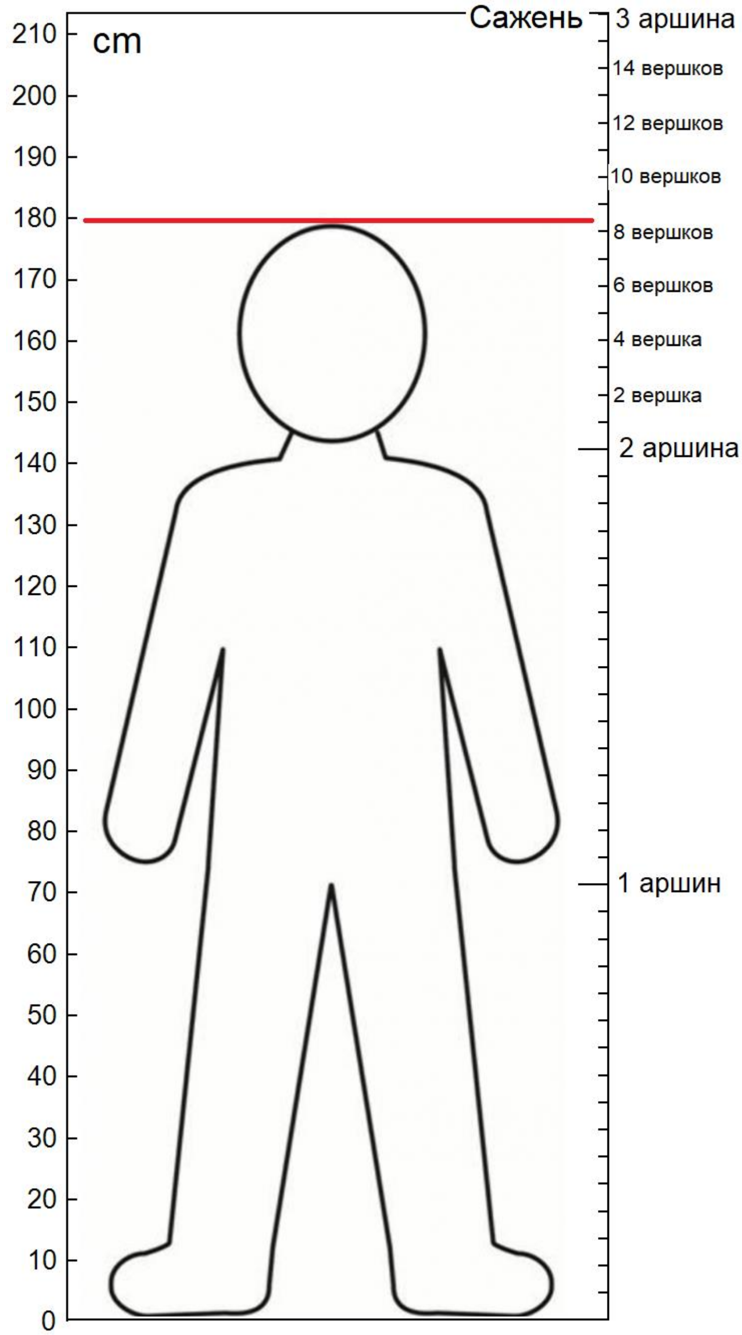


Рис. 6.9. Рост взрослого мужчины

А давайте всех студентов-юношей, показанных на рис. 6.1, расставим в семь рядов с интервалом в один вершок, написав небольшую программу – см. рис. 6.10. Мы получим некий цифровой двойник, ситуации, показанной на рис. 6.7.

```
Hist :=  | Hist := [0 0 0 0 0 0 0]^T
         | for i ∈ [1..n]
         |   if Рост_i < 5.5 вершок
         |     Hist_1 := Hist_1 + 1
         |   if 5.5 вершок ≤ Рост_i < 6.5 вершок
         |     Hist_2 := Hist_2 + 1
         |   if 6.5 вершок ≤ Рост_i < 7.5 вершок
         |     Hist_3 := Hist_3 + 1
         |   if 7.5 вершок ≤ Рост_i < 8.5 вершок
         |     Hist_4 := Hist_4 + 1
         |   if 8.5 вершок ≤ Рост_i < 9.5 вершок
         |     Hist_5 := Hist_5 + 1
         |   if 9.5 вершок ≤ Рост_i < 10.5 вершок
         |     Hist_6 := Hist_6 + 1
         |   if Рост_i ≥ 10.5 вершок
         |     Hist_7 := Hist_7 + 1
         | augment([5 6 7 8 9 10 11]^T, Hist)
```

le

lele

lelt

lelt("1:переменная", "2:число", "3:число") - Boolean comparison. Check if: "2:number" ≤ "1:variable" < "3:number".

Рис. 6.10. Программа для построения гистограммы

Программа на рис. 6.10 возвращает матрицу с семью строками и двумя столбцами, которая на рис. 6.11 с помощью программы-функции с именем *drow_his*, используется для построения гистограммы – столбчатой диаграммы.

$$draw\_hist(Hist) := \left|\begin{array}{l} \delta := \dfrac{Hist_{2\,1} - Hist_{1\,1}}{2} \\ xy := 0 \\ \text{for } \ i \in [1..rows(Hist)] \\ \quad \left|\begin{array}{l} xy1 := augment\left(\begin{bmatrix} Hist_{i\,1} - \delta \\ Hist_{i\,1} - \delta \\ Hist_{i\,1} + \delta \\ Hist_{i\,1} + \delta \end{bmatrix}, \begin{bmatrix} 0 \\ Hist_{i\,2} \\ Hist_{i\,2} \\ 0 \end{bmatrix}\right) \\ xy := stack(xy, xy1) \end{array}\right. \\ xy \end{array}\right.$$

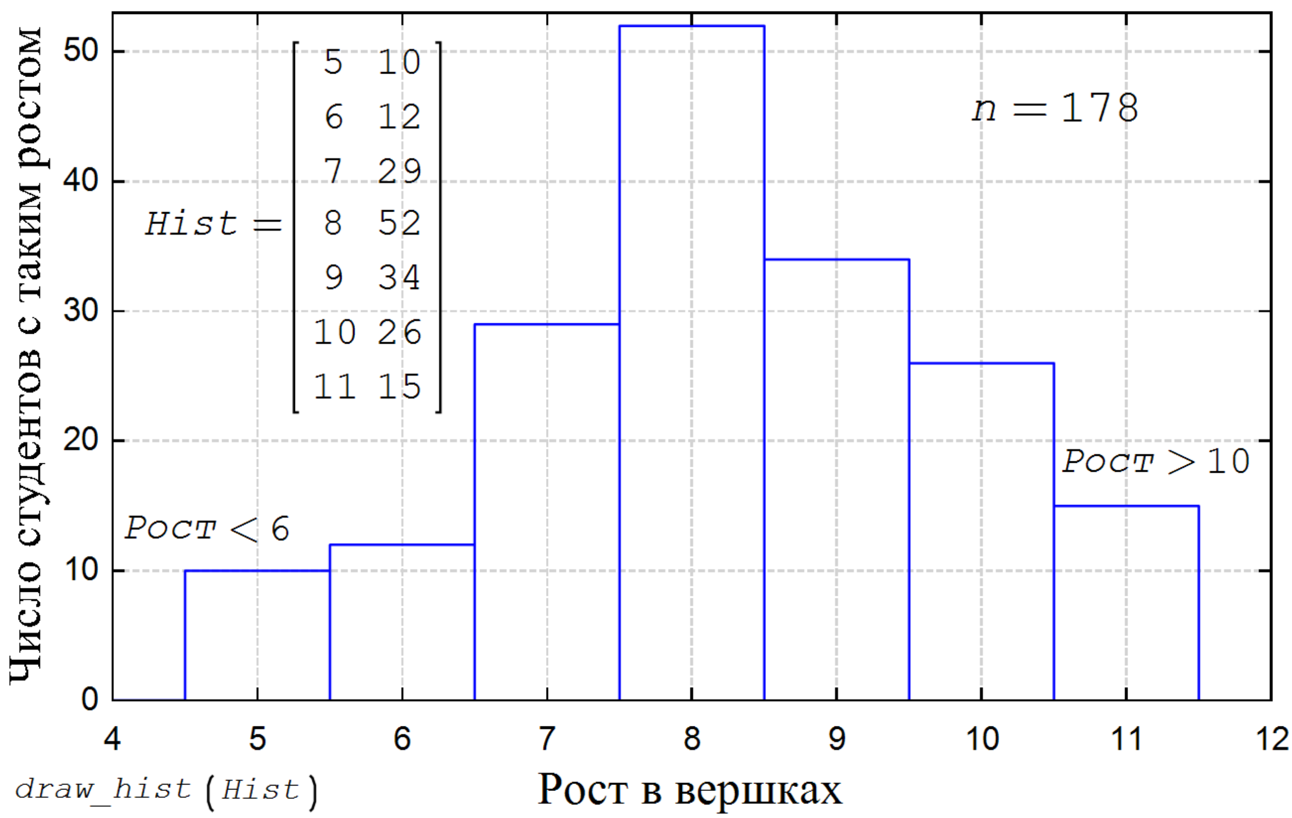


Рис. 6.11. Гистограмма роста студентов

Гистограмма на рис. 6.11 позволяет нам при измерении роста взрослого мужчины перейти от числовых к лингвистическим константам, что является предметом раздела математики под названием «Теория нечетких множеств». Человек может быть очень низким; низким; скорее низким, чем высоким; среднего роста; скорее высоким, чем низким; высоким и, наконец, очень высоким. В этом отношении русский вершок намного удобнее, чем европейский сантиметр или англо-американский дюйм.

Мы отметили, что матрица с именем `Data`, показанная наверху рис. 6.2, – это по своей сути база данных с двумя полями и 20 записями.

Одно из главных действий при работе с базой данных – это сортировка. В среде SMath для этого предусмотрены соответствующие функции, работа одной из которых (`rsort`) показана на рис. 6.12: наших студентов расставили по росту (первая строка (row) матрицы) и по весу (вторая строка).

$$Data := \begin{bmatrix} 189 & 177 & 179 & 183 & 180 & 185 & 178 & 186 & 176 & 185 & 184 & 180 & 180 & 188 & 183 & 195 & 194 & 188 & 182 & 195 \\ 90 & 80 & 73 & 73 & 85 & 83 & 75 & 75 & 77 & 73 & 76 & 74 & 68 & 81 & 79 & 88 & 75 & 89 & 85 & 80 \end{bmatrix}$$

$$\mathrm{rsort}(Data, 1) = \begin{bmatrix} 176 & 177 & 178 & 179 & 180 & 180 & 180 & 182 & 183 & 183 & 184 & 185 & 185 & 186 & 188 & 188 & 189 & 194 & 195 & 195 \\ 77 & 80 & 75 & 73 & 85 & 68 & 74 & 85 & 73 & 79 & 76 & 83 & 73 & 75 & 81 & 89 & 90 & 75 & 88 & 80 \end{bmatrix}$$

$$\mathrm{rsort}(Data, 2) = \begin{bmatrix} 180 & 179 & 183 & 185 & 180 & 194 & 178 & 186 & 184 & 176 & 183 & 177 & 195 & 188 & 185 & 180 & 182 & 195 & 188 & 189 \\ 68 & 73 & 73 & 73 & 74 & 75 & 75 & 75 & 76 & 77 & 79 & 80 & 80 & 81 & 83 & 85 & 85 & 88 & 89 & 90 \end{bmatrix}$$

Рис. 6.12. Встроенная функция сортировки матрицы

Функция `csort` ведет сортировку не по рядам, а по столбцам матрицы (см. рис. 9.11, занятие 9), а функция `sort` сортирует вектор – матрицу с одним столбцом.

А какой алгоритм заложен в перечисленные функции? Разработано большое количество алгоритмов для этой работы, совершенствование которых сводилось в основном к повышению их быстродействия. А давайте в ключе занятия 2 (возрождение/ренессанс простых алгоритмов) соорудим собственную простую и понятную программу сортировки вектора. И, главное, покажем одну особенность пакета SMath, о которой не знают даже опытные пользователи.

При программировании иногда нужно менять местами значения в двух переменных. Это обычно делается *последовательным* перемещением значения между тремя переменными по схеме треугольника – см. верхнюю половину рис. 6.13.

Последовательное вычисление

$a := 1 \qquad b := 2$

$c := a \qquad a := b \qquad b := c$

$a = 2 \qquad b = 1$

Параллельное вычисление

$[a\ \ b] := [b\ \ a]$

$a = 1 \qquad b = 2$

Рис. 6.13. Два способа обмена значениями двух переменных

Но если переменные объединить в массив (вторая половина рис. 6.13), то операторы присваивания (обмена) будут работать как бы параллельно. Эта особенность используется в программе сортировки на рис. 6.14.

```
MySort(v) := | Флаг := 1
             | while Флаг
             |   | Флаг := 0
             |   | for i ∈ [2..length(v)]
             |   |   if v[i-1] > v[i]
             |   |     [v[i-1]  v[i]  Флаг] := [v[i]  v[i-1]  1]
             | v
```

$v := [1\ 8\ 2\ -6\ 3\ 5\ 4] \qquad v := MySort(v) = [-6\ 1\ 2\ 3\ 4\ 5\ 8]$

Рис. 6.14. Простейшая программа сортировки

6.2. Гистограмма математиков

Когда-то давно два студента автора никак не могли сдать зачеты по двум дисциплинам первого курса «Информационные технологии» и «Высшая математика». Было решено испытать не их ум, а их руки и заставить этих студентов «освоить» математику с информатикой хотя бы так. Первый студент открывает страницы Википедии[1] с информацией об умерших великих и выдающихся[2] математиках и диктует второму студенту, сколько лет этот математик прожил.

[1] А там есть русскоязычная статья с алфавитным списком математиков и ссылками на их страницы в Википедии.

[2] Формальным критерием того, что данный математик является великим или просто выдающимся, был тот факт, что о нем написано более чем две страницы текста не только в сегменте Википедии родного языка этого математика, но и в сегментах таких языков: английский, арабский, испанский, китайский, русский, французский, итальянский и немецкий.

Второй студент заносит эту информацию в расчётный лист SMath. По этим данным строится гистограмма – см. рис. 6.15.

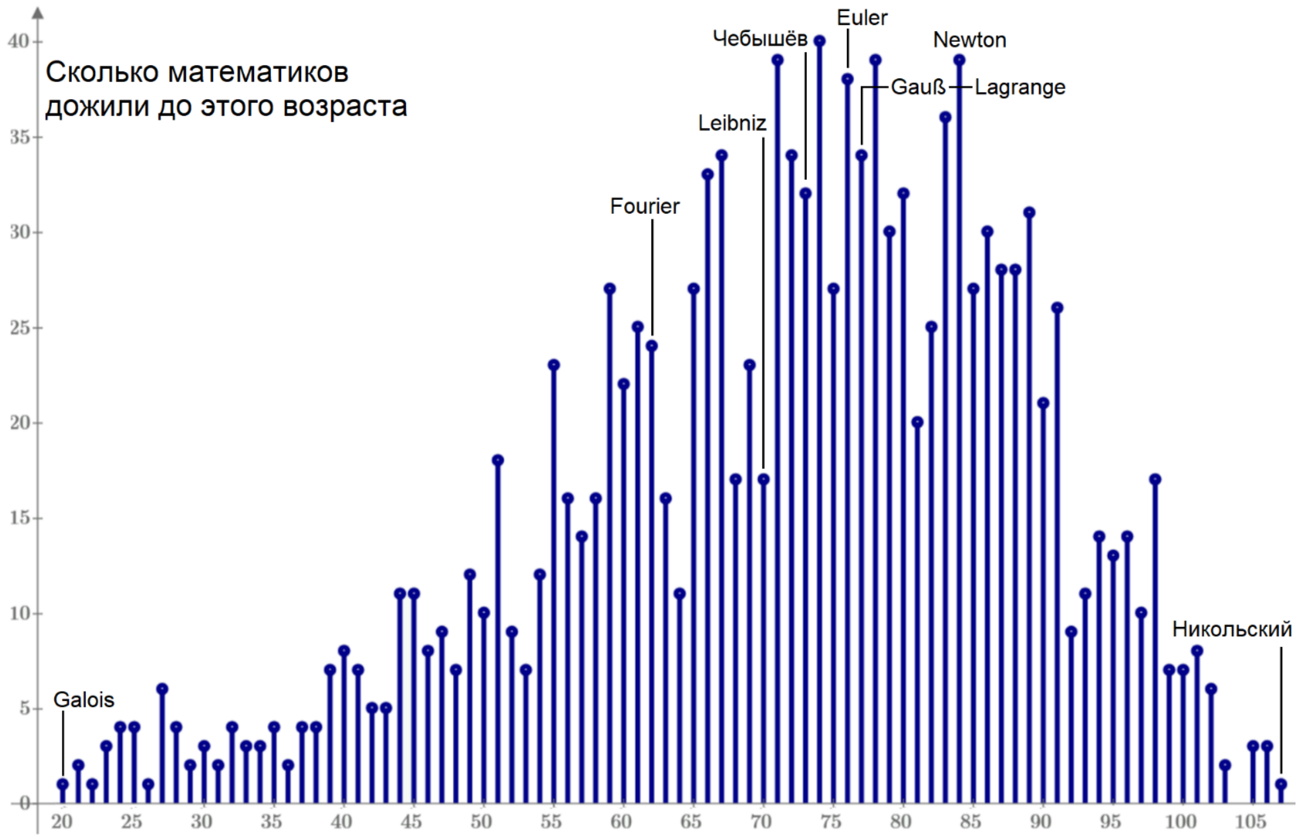


Рис. 6.15. Гистограмма математиков

Гистограмма охватила почти две тысячи великих и выдающихся математиков из Википедии. Ось абсцисс – сколько они прожили лет, а ось ординат – это сколько математиков дожили до этого возраста [5].

Вернемся к регрессионному анализу.

На рисунке 6.16 показан авторский сайт интернете, где по возрасту легкового автомобиля и его пробегу, можно оценить, сколько он потерял в цене. Это задача более сложная, более интересная и более практичная.

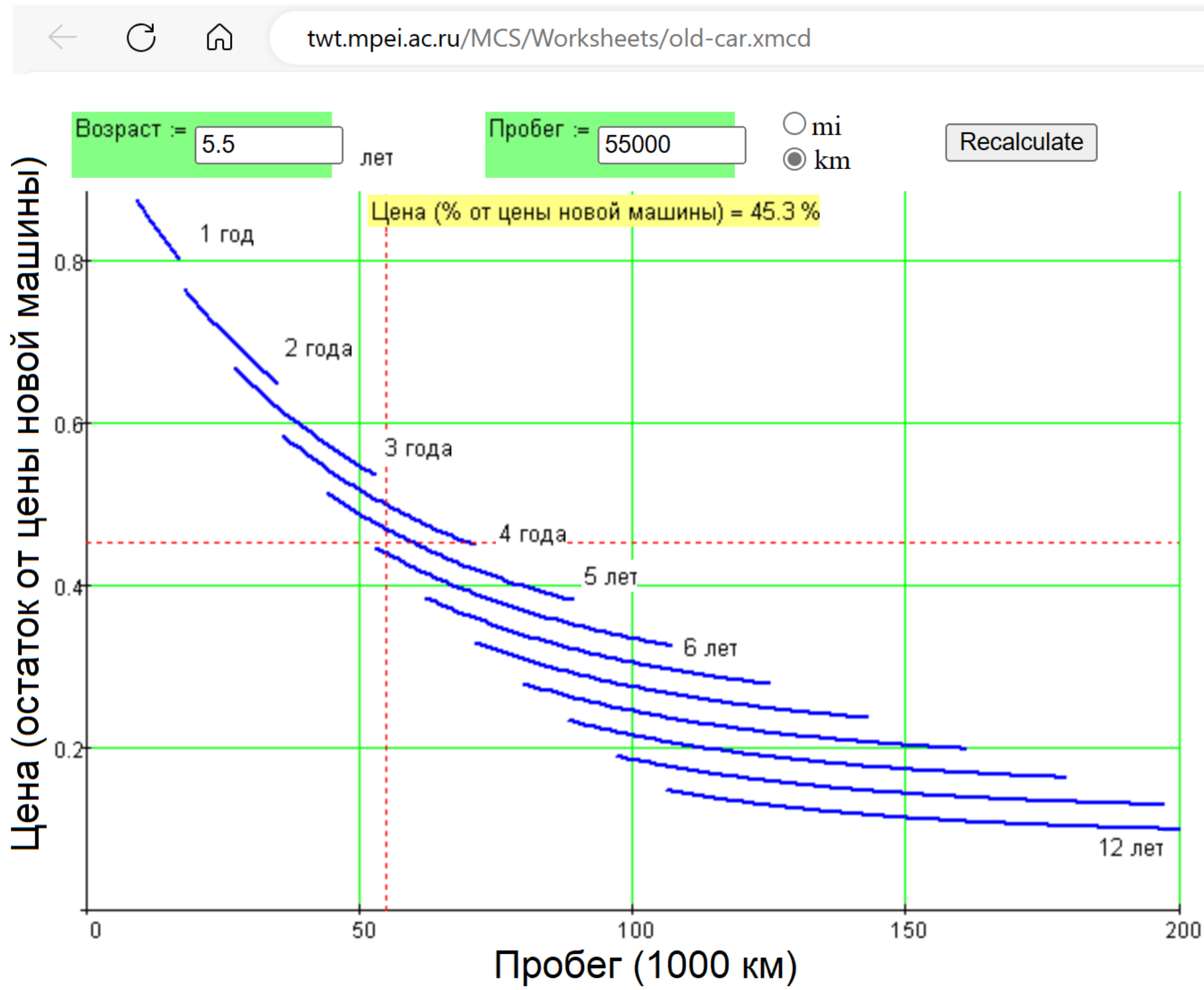


Рис. 6.16. Аппроксимация данных по стоимости подержанного автомобиля

Автор три раза покупал себе новую машину по схеме trade-in. Трейд-ин — это обмен старой машины на другую с доплатой. Сделка проходит так: клиент приезжает в салон на своей машине, там её оценивают и, если клиента устраивает цена, принимают авто в счёт частичной оплаты нового. Потом дилер перепродаст машину клиента и вернёт себе деньги. Увы, у нас в стране довольно часто в такой ситуации слово «обмен» заменяют на слово «обман». Работники автосалонов все три раза пытались занизить цену сдаваемой старой машины. Приходилось спорить и показывать на смартфоне сайт, отображенный на рис. 6.16. Это помогало поднять цену сдаваемой машины.

Данная задача подробно описана на форуме пользователей SMath по адресу https://en.smath.com/forum/yaf_postst23733_Why-I-have-errors.aspx.

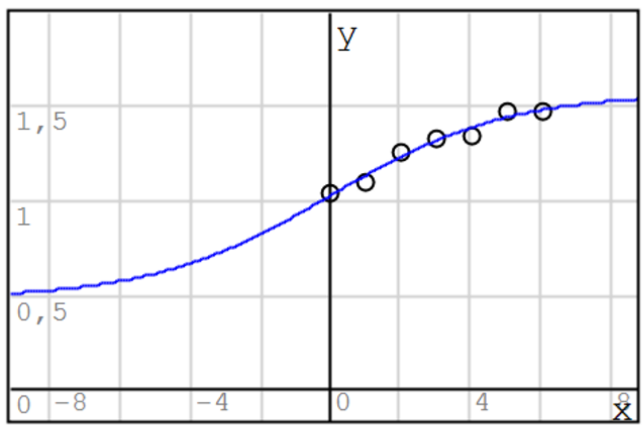
Для нелинейного регрессионного анализа можно использовать функцию *Fit* из плагина Maxima для SMath – см. рис. 6.17. Аргументы этой функции включают матрицу данных, список с названиями столбцов матрицы данных, статистическую модель, включающую данные величины и параметры подгонки, список с именами параметров и список с предполагаемыми значениями параметров. Функция возвращает список уравнений вида «имя = значение». Эти уравнения преобразуются в оператор присваивания с использованием функции *Assign* также из плагина Maxima. В данном примере мы сгенерировали измерения длины маятника (см. занятие 5) с использованием функции `L(F)` и добавили некоторую имитированную экспериментальную ошибку. При расчёте можно просто ограничиться работой с векторами *F* и *L*, объединив их в матрицу *D*. См. также рисунок 5.10 (занятие 5), где показано практическое применение функции *Fit* для расчёта маятника на эластичной подвеске.

$$F := \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix} \qquad L := \overrightarrow{\frac{L(F\ \mathrm{N})}{\mathrm{m}}} = \begin{bmatrix} 1 \\ 1,12 \\ 1,23 \\ 1,32 \\ 1,38 \\ 1,42 \\ 1,45 \end{bmatrix} \qquad \Delta := \frac{\mathrm{Random}(7) - 0,5}{10} = \begin{bmatrix} 0,0335 \\ -0,0312 \\ 0,0113 \\ 0,00801 \\ -0,0426 \\ 0,0424 \\ 0,00481 \end{bmatrix}$$

$$l(f) := a \cdot \left(\frac{1}{1 + \mathbf{e}^{-c \cdot f}} + b \right) \qquad D := \mathrm{augment}(F; L + \Delta)$$

$$L := \mathrm{Fit}\left(D; \begin{cases} f \\ l \end{cases}; l = l(f); \begin{cases} a \\ b \\ c \end{cases}; \begin{cases} 1 \\ 1 \\ 1 \end{cases} \right) = \begin{cases} a = 1,0842031409351915 \\ b = 0,439584391663257 5 \\ c = 0,3976130029475916 \end{cases}$$

$$\mathrm{Assign}(L) = \begin{cases} 1,08 \\ 0,44 \\ 0,398 \end{cases}$$



$$\begin{cases} l(x) \\ \mathrm{augment}(D; \text{"o"}) \end{cases}$$

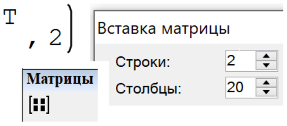
Рис. 6.17. Аппроксимация с помощью плагина Maxima

В начале этого занятия мы сказали, что конкурсы красоты могут заканчиваться… судебными тяжбами. Вернемся к рисунку 6.2. Студент с ростом 183 см и весом 79 кг (второе место на конкурсе) может возразить в том плане, что неправильно считались отклонения – они учитывались по вертикали (см. п.1 на рис. 6.18), а нужно учитывать истинные (ортогональные) такие расстояния по перпендикуляру (п. 2) – по наикратчайшему пути от точки до прямой (см. https://www.dmitrymakarov.ru/linear-algebra/lstsq). В расчете, показанном на рис. 6.18, вводится целевая функция с именем CO (сумма отклонений) для расчета суммы расстояний от каждой точки до прямой, описываемой уравнением $a \cdot x + b \cdot y + c = 0$. На сайте пользователей SMath https://en.smath.com/forum/yaf_postsm85477_SMath-vs-Mathcad.aspx было подсказано, как найти значения коэффициентов этого уравнения, при которых функция CO примет минимальное значение. Эти коэффициенты введены в задаче на рис. 6.18, где на графике прочерчены две прямые линии – красная (п. 1) отображает стандартный метод регрессионного линейного анализа (рис. 6.2), а зелёная (п. 2) – наш модифицированный метод.

В нижней части рис. 6.18 показаны три сечения четырехмерной поверхности функции CO вблизи точки минимума – точки решения нашей задачи. Кривые на графиках ломаные. Это является следствием того, то в функции CO записан оператор абсолютного значения, а не возведения во вторую степень – см. гладкие сечения на рис. 6.6, где отображалась функция CKO со второй степенью, а не с модулем как на рис. 6.18.

$$Data := \begin{bmatrix} 189 & 177 & 179 & 183 & 180 & 185 & 178 & 186 & 176 & 185 & 184 & 180 & 180 & 188 & 183 & 195 & 194 & 188 & 182 & 195 \\ 90 & 80 & 73 & 73 & 85 & 83 & 75 & 75 & 77 & 73 & 76 & 74 & 68 & 81 & 79 & 88 & 75 & 89 & 85 & 80 \end{bmatrix}$$

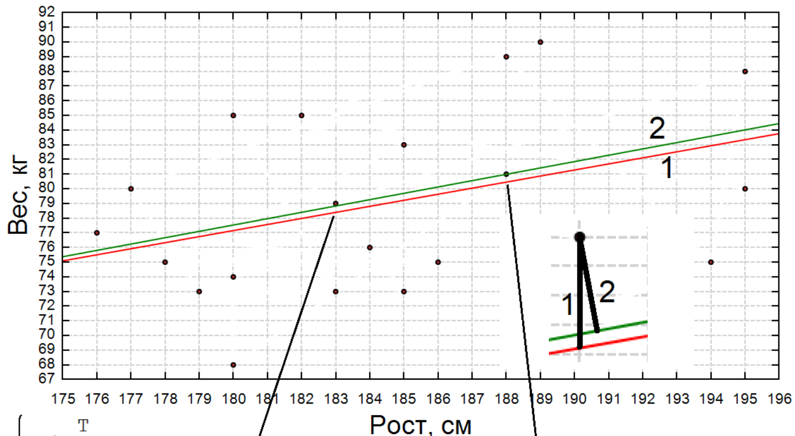
$n := \mathrm{cols}(Data) = 20$ $\quad$ $Рост := \mathrm{col}(Data^{T}, 1)$ $\quad$ $Вес := \mathrm{col}(Data^{T}, 2)$



$a := \mathrm{Intercept}(Рост, Вес) = 2.833$ $\quad$ $b := \mathrm{Slope}(Рост, Вес) = 0.4129$

$$вес(рост) := a + b \cdot рост$$

$$CO(a, b, c) := \sum_{i=1}^{n} \frac{\left| a \cdot Рост_{i} + b \cdot Вес_{i} + c \right|}{\sqrt{a^{2} + b^{2}}}$$ $\quad$ $a_1 := -1.5206$ $\quad$ $b_1 := 3.5175$ $\quad$ $c_1 := 0.9523$

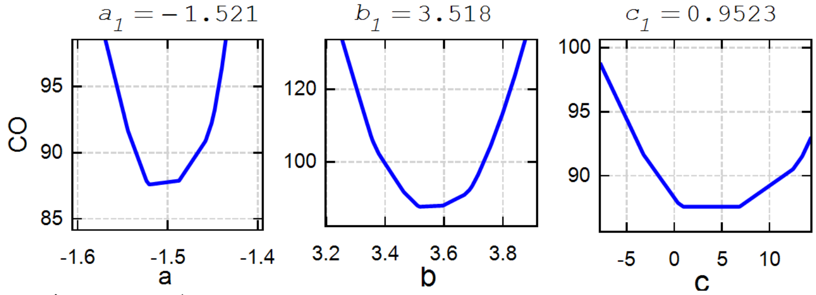


$$\begin{cases} Data^{T} \\ вес(рост) \\ a_1 \cdot x + b_1 \cdot y + c_1 \end{cases}$$

$$\frac{\left| a_1 \cdot 188 + b_1 \cdot 81 + c_1 \right|}{\sqrt{a_1^{2} + b_1^{2}}} = 0.0007829$$

$$\frac{\left| a_1 \cdot 183 + b_1 \cdot 79 + c_1 \right|}{\sqrt{a_1^{2} + b_1^{2}}} = 0.1474$$

$a_1 = -1.521$ $\quad$ $b_1 = 3.518$ $\quad$ $c_1 = 0.9523$



$\mathrm{LS}(x, b_1, c_1)$ $\quad$ $\mathrm{LS}(a_1, x, c_1)$ $\quad$ $\mathrm{LS}(a_1, b_1, x)$

Рис. 6.18. Модифицированный метод линейного регрессионного анализа

Вице-мистеру конкурса красоты не удалось оспорить его результаты. Более того, результат победителя конкурса красоты стала ещё убедительней – см. подсчёты двух расстояний внизу рисунка 6.18. Но вице-мистер может попытаться апеллировать к медиан-медианной регрессии. Читатель, реализуй этот метод в среде SMath, опираясь на описания в интернете!

Победители победителями, но почему же при линейном регрессионном анализе используют сумму квадратов «вертикальных» отклонений, а не сумму истинных расстояний. Отклонения могут иметь разный знак, поэтому и их возводят в квадрат. Расстояния же всегда имеют положительное значение. Дело в том, что любой метод поиска тренда в статистических выборках весьма условен и приводит к довольно спорным результатам. Недаром говорят, что есть большая ложь, маленькая ложь и… статистика. Причина тут в том, что учёт истинного отклонения был практически нереализуем в докомпьютерную эру.

А теперь поговорим об ***интерполяции*** – о случае, когда линия проходит не вблизи точек (см. рис. 6.2 и 6.17), а через сами точки.

На рисунке 6.19 показан простейший вид интерполяции – кусочно-линейная интерполяция, реализуемая в среде SMath встроенной функцией с именем `linterp`. Данные в матрице Data произвольные.

$$Data := \begin{bmatrix} 1 & 2 & 3 & 4.5 & 5 & 6.1 & 7 & 8 & 8.9 & 10 & 11 & 12 \\ 4 & 5 & 7 & 9 & 6 & 7 & 11 & 24 & 34 & 40 & 37 & 47 \end{bmatrix} \qquad n := \mathrm{cols}(Data) = 12$$

$$X := \mathrm{row}(Data, 1)^{T} \qquad Y := \mathrm{row}(Data, 2)^{T} \qquad y(x) := \mathrm{linterp}(X, Y, x)$$

$$x := \left[\min(X), \min(X) + \frac{\max(X) - \min(X)}{1000} .. \max(X) \right]$$

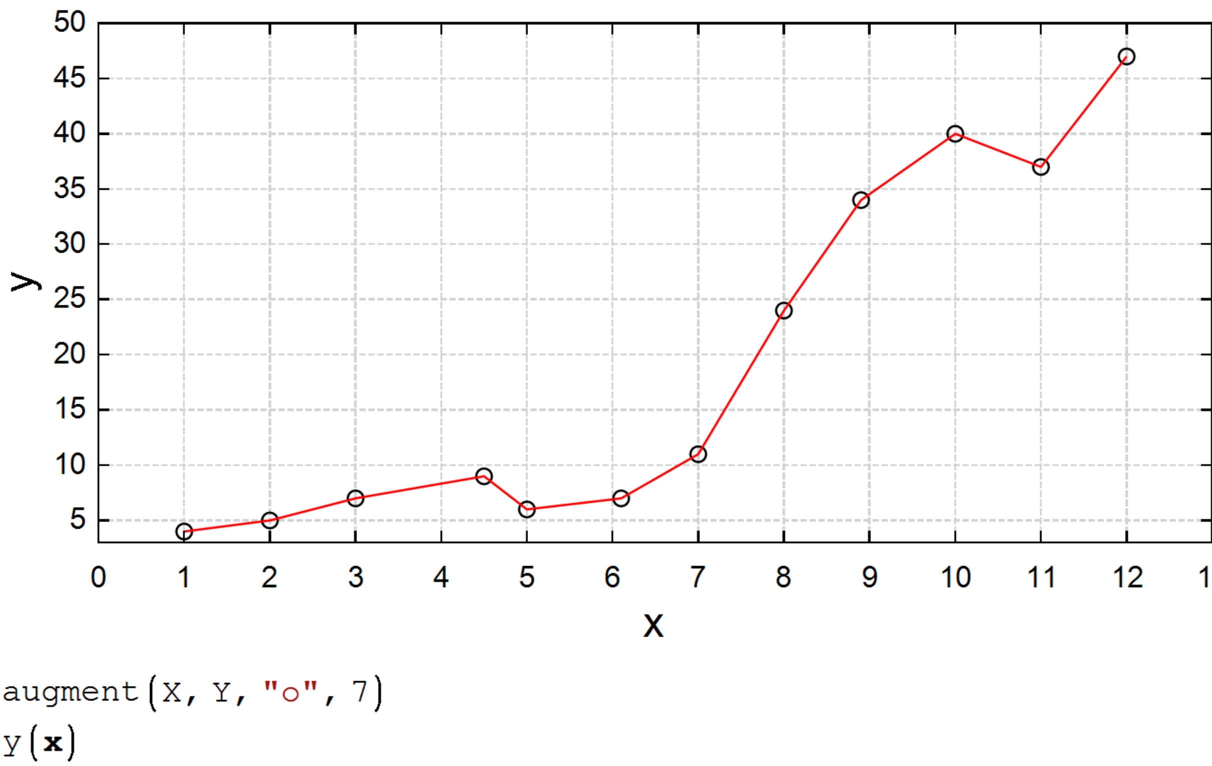


Рис. 6.19. Кусочно-линейная интерполяция

Если вернуться к рисунку 6.4, то можно вспомнить и об интерполяции сплайнами – кубическими сплайнами, когда через две соседние точки проводится кривая полинома третьей степени, но так, чтобы значения вторых производных соседних кубических парабол в узловых точках были равны.

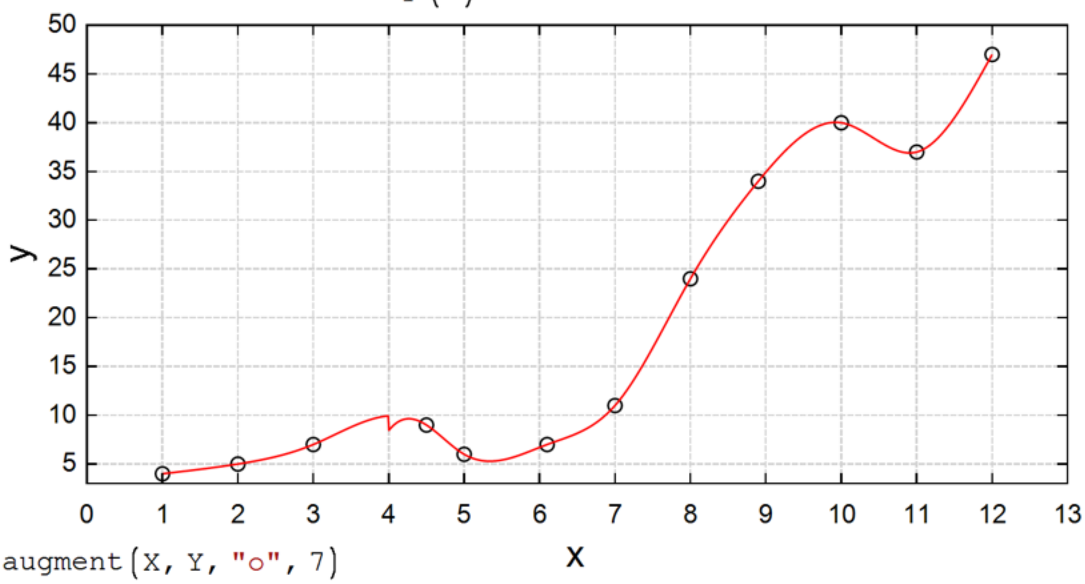
На рисунке 6.20 показана процедура проведения сплайн-интерполяции по точкам, введенным в расчёт на рис. 6.19.

$k := \mathrm{lspline}(X, Y)$ Функция из дополнения Mathcad Tools

$k^{T} = [\,0\ 3\ 0\ 0\ 0.663\ 3.35\ -14.3\ 16.1\ -2.7\ 15.9\ -6.62\ -2.28\ -17.1\ 23.8\ 0\,]$

$y(x) := \mathrm{interp}(k, X, Y, x)$ Функция из дополнения Mathcad Tools

$y(6) = 6.6959$



$\begin{cases} \mathrm{augment}(X, Y, \text{"o"}, 7) \\ y(x) \end{cases}$

Рис. 6.20. Сплайн-интерполяция

Функция `lspline` возвращает вектор, первые три элемента которого – это служебная информация, а остальные элементы – это значения вторых производных искомой интерполирующей функции в узловых точках, коих у нас двенадцать. На рисунке 6.20 используется встроенная функция `interp`, которая с опорой на функцию `lspline` проводит сплайн-интерполяцию – генерирует нужную функцию. Между третьей и четвертой точками виден некий дефект. Давайте исправим его и заодно поймем суть сплайн-интерполяции, а также создадим некое произведение изобразительного искусства.

На рисунке 6.20 показан весь набор кубических парабол, плавно переходящих одна в одну. Эти кривые третьего порядка можно нарисовать отдельно, указывая номер точки. Такой расчёт показан на рис. 6.21 для третьей точки, по которой функция `lspline` выдала небольшой дефект – см. рис. 6.20.

Для этого решается система четырех алгебраических уравнений с четырьмя неизвестными a, b, c и d, формирующих функции пользователя cp (кубический полином) и cp'' (его вторая производная). На рисунке 6.21 такой расчёт сделан для интервала от третьей до четвертой точки ($j = 3$), где функция `lspline` имела дефект.

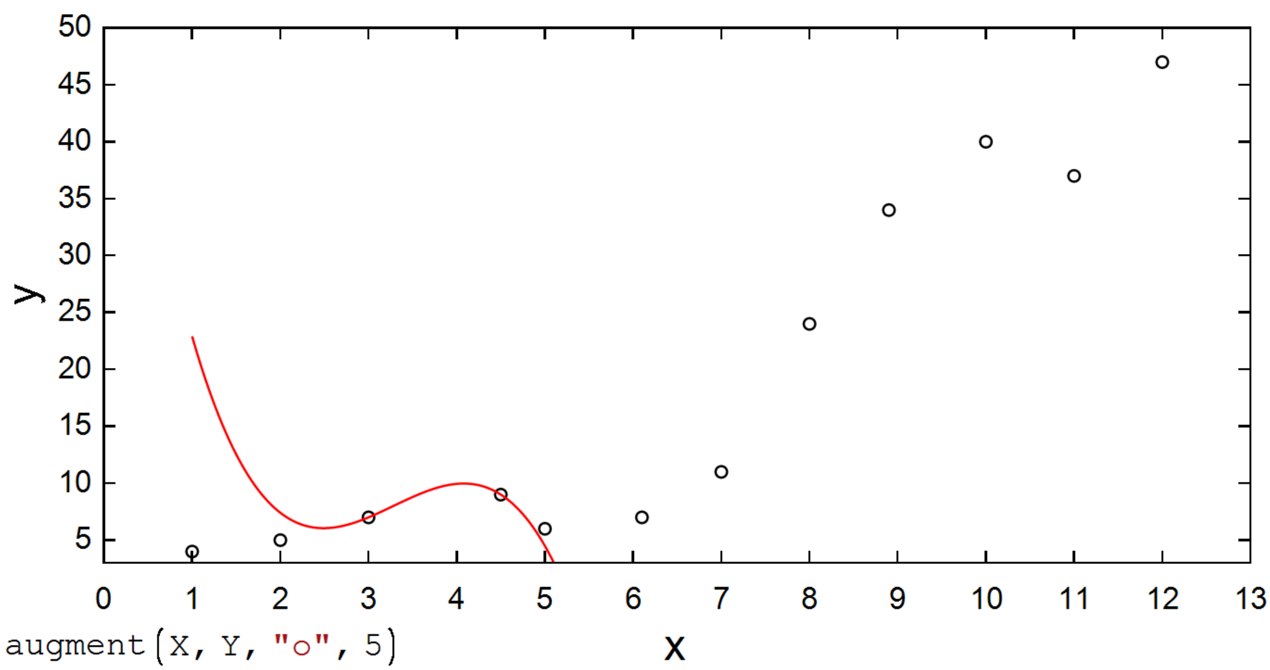
$$cp(x, a, b, c, d) := a + b \cdot x + c \cdot x^2 + d \cdot x^3$$

$$cp''(x, c, d) := 2 \cdot c + 6 \cdot d \cdot x$$

3 See the plot bellow

j

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} := \left(\text{roots} \left(\begin{bmatrix} cp\left(X_j, a, b, c, d\right) = Y_j \\ cp''\left(X_j, c, d\right) = k_{j+3} \\ cp\left(X_{j+1}, a, b, c, d\right) = Y_{j+1} \\ cp''\left(X_{j+1}, c, d\right) = k_{j+3+1} \end{bmatrix}, \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \right) \right) = \begin{bmatrix} 65.2426 \\ -59.6732 \\ 19.2924 \\ -1.9576 \end{bmatrix}$$



$$\begin{cases} \text{augment}(X, Y, \text{"o"}, 5) \\ \text{cp}(\mathbf{x}, a, b, c, d) \end{cases}$$

Рис. 6.21. Одна из кубических парабол, полученная при сплайн-интерполяции

Если менять значение переменной j от 1 до 12 (это на рис. 6.21 удобно и быстро делается с помощью специального элемента управления, который был описан на занятии 2 – см. рис. 2.3), то можно построить все 12 кубических парабол – см. рис. 6.22.

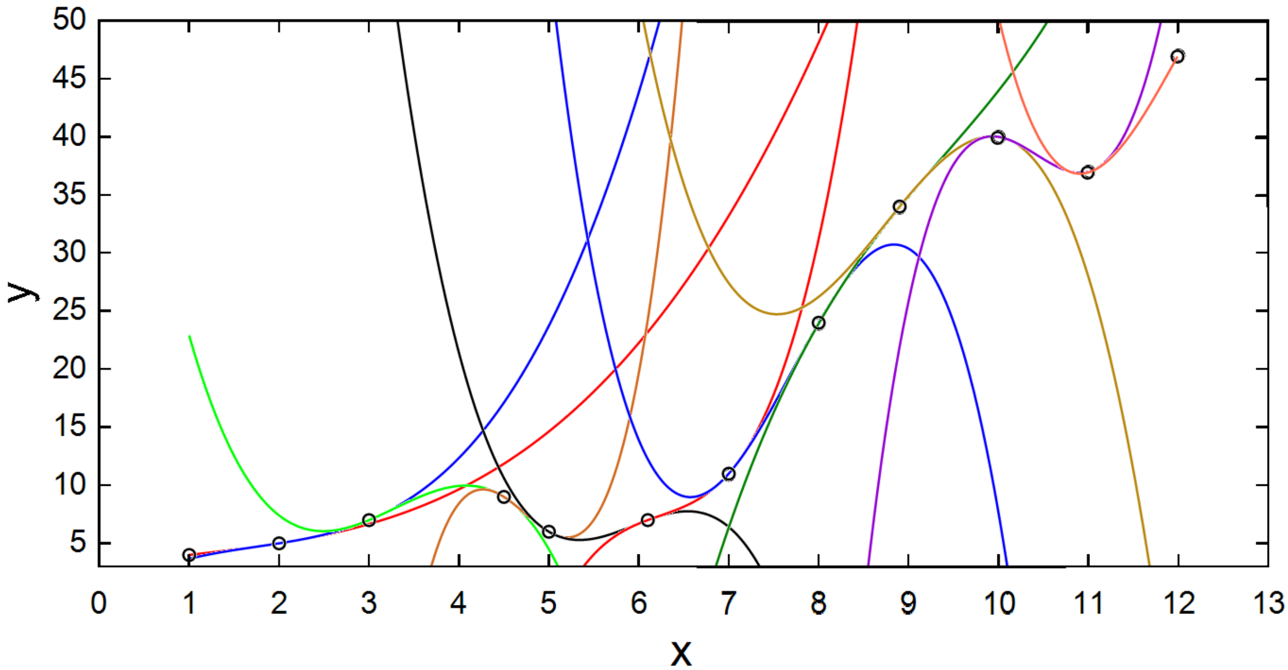


Рис. 6.22. Все кубические параболы, полученные при сплайн-интерполяции

В рисунке 6.22 показано, как можно стереть лишнее, дорисовать недостающее и получить некое произведение изобразительного искусства с четким математическим смыслом – рис. 6.23. Из таких ветвей можно свить венок и увенчать им победителя конкурса красоты, с которого мы начали это занятие.

Рис. 6.23. Ветка дерева с плодами сплайн-интерполяции

6.3. От таблицы к функции

А как получить интерполяционную функцию двух аргументов?

В одном старом польском фильме «Гангстеры и филантропы»[3] есть такой сюжет. Одного человека уволили с работы из-за того, что он по халатности устроил маленький взрыв по месту службы в химической лаборатории. Этот бедолага заходит в ресторан и с горя заказывает себе стакан водки и закуску – селедку[4] с хлебом (см. рис. 6.24). В печальных раздумьях этот химик машинально вытаскивает из нагрудного кармана пиджака ареометр и начинает им помешивать водку в стакане. Персонал думает, что это контролёр, проверяющий крепость напитка (не разбавляют ли водку водой[5]), и дает ему взятку. Наш безработный

[3] Его можно посмотреть на YouTube. Во времена соцлагеря только польские фильмы были более-менее приемлемы для просмотра. Вспомним хотя бы легендарный фильм "Пепел и алмаз". Остальное «социалистическое кино» было откровенной мурой.

[4] На рисунке 6.24 видно, что к селедке подали белый, а не черный хлеб – это Польша, а не Россия. Да и стакан не граненый!

[5] В советском культовом фильме с прекрасными актерами "Дайте жалобную книгу!" есть такая сцена. Официантка подходит к буфету, заказывает ещё одну бутылку водки и жалуется, что её клиенты уже здорово набрались. Буфетчица под прилавком разбавляет водку водой и дает официантке: "На, вот, отнеси им вот это!".

химик догадывается, что к чему и начинает так «химичить» и в других ресторанах Варшавы. Отечественный самый знаменитый сюжет на эту тему описан Гоголем в пьесе «Ревизор».

Рис. 6.24. Кадр из фильма «Гангстеры и филантропы»

Ареометр – это один из подручных измерительных приборов химиков-аналитиков. Неслучайно он оказался в кармане нашего химика. Прибор представляет собой стеклянный поплавок со шкалой погружения. Ареометром можно измерить плотность раствора, а затем по ней определить его концентрацию. Автовладельцы старшего поколения сразу вспомнят свинцовые аккумуляторы с залитым в них раствором серной кислоты. Её концентрацию нужно было периодически проверять специальным ареометром и подливать в аккумулятор при необходимости дистиллированную воду[6], которая со временем испарялась, что требовало новой проверки. Если шкалу ареометра разметить в градусах (процентах) спиртных напитков, то он будет называться градусником, пардон, спиртометром. Градусник – это совсем другой измерительный прибор, который тоже нужно опустить в раствор. Только мольная

[6] В те времена первый автор был аспирантом кафедры Технологии воды и топлива МЭИ, которая снабжала автомобилистов всего института дистиллированной водой.

концентрация не зависит от температуры (отношение количества растворенного вещества к массе растворителя).

А давайте оценим теплофизические параметры водки в среде SMath с подгруженным к ней пакетом функций CoolProp. Это будет неким поучительным вычислительным экспериментом. Отечественную математическую программу SMath можно свободно за пару минут скачать с сайта www.smath.com, а затем опять же за пару минут дооборудовать её функциями, возвращающими теплофизические свойства жидкостей, газов и их смесей. Это делается через цепочку команд Сервис/Опции/Дополнения/Галерея онлайн/CoolProp Wrapper. Приемы работы в среде SMath описаны в учебном пособии http://twt.mpei.ac.ru/ochkov/EC-SMath.pdf. Сведения же о пакете CoolProp можно найти на сайте www.coolprop.org. Окончание адреса org означает, что это некоммерческая организация. Имена сайтов с программами, которые нужно приобретать за деньги[7], обычно оканчиваются на com – коммерческий сайт. Платная версия программы CoolProp называется RefProp. Её можно скачать с сайта https://www.nist.gov/srd/refprop, принадлежащего Национальному институту стандартов и технологий (NIST – США). Cool и Ref – это начала английских слов Coolant (охладитель) и Refrigerant (хладогент). Prop же – это начало слова Property – свойство. Термин Cool оказался удачным и потому, что на молодежном сленге это английское слово имеет и такое значение – «клевый»! Но компьютерный переводчик выдал более трех десятков вариантов перевода на русский слова Cool. Посмотрим, насколько CoolProp оказался «клевым» по отношению к водке.

Итак, давайте, как сейчас принято говорить, создадим цифрового двойника ситуации, зафиксированной на рис. 6.24.

На рисунке 6.25 показано, как в расчет вводятся значения температуры и давления, при которых далее будут определяться теплофизические параметры воды и безводного этилового спирта.

На рисунках 6.25 и 6.26 показан вызов в среде SMath двух функций из пакета CoolProp Wrapper – функции `CoolProp_get_fluid_param_string` с двумя аргументами и функции `CoolProp_Props` с шестью аргументами. Первая функция вернула нам

[7] Анекдот из будущих времен ослабления санкций.
– *Теперь фотошоп снова можно купить на сайте разработчика.*
– *Что!? Купить?!*

химическую формулу, а вторая плотность ("D"), молярную массу ("M"), температуру кипения при атмосферном давлении ("T"), давление кипения (упругость паров) при нормальной температуре ("P"), вязкость ("V" – viscosity), массовую удельную теплоемкость ("Cpmass") и теплопроводность ("conductivity"). Полный список параметров и веществ можно найти на вышеотмеченном сайте www.coolprop.org. Единица, проставленная в качестве пятого аргумента функции `CoolProp_Props` при расчете температуры и давления кипения на линии насыщения, указывает на то, что мы имеем дело с сухим насыщенным паром. Если единицу заменить на нуль, то это будет жидкость на линии насыщения. Вместо единицы тут можно поставить любое число в диапазоне от нуля до единицы. На температуру и давление кипения это не будет влиять, чего не скажешь при определении других параметров – удельной энтальпии, например.

$T := 20\ °C \qquad p := 1\ \text{атм}$

CoolProp_get_fluid_param_string("H2O"; "formula") = "H_{2}O_{1}"

$\rho_{H2O} :=$ CoolProp_Props("D"; "P"; p; "T"; T; "H2O") = $0{,}9982\ \dfrac{\text{г}}{\text{см}^3}$

CoolProp_Props("M"; "P"; p; "T"; T; "H2O") = $18{,}02\ \dfrac{\text{г}}{\text{моль}}$

CoolProp_Props("T"; "P"; p; "Q"; 1; "H2O") = $100\ °C$

CoolProp_Props("P"; "T"; T; "Q"; 1; "H2O") = $17{,}55$ мм рт.ст.

CoolProp_Props("V"; "P"; p; "T"; T; "H2O") = $1{,}002\ \dfrac{\text{Н мм}}{\frac{\text{м}}{\text{с}}\ \text{м}^2}$

CoolProp_Props("Cpmass"; "P"; p; "T"; T; "H2O") = $0{,}9993\ \dfrac{\text{кал}}{\text{г К}}$

CoolProp_Props("conductivity"; "P"; p; "T"; T; "H2O") = $598\ \dfrac{\text{Вт мм}}{\text{м}^2\ \text{К}}$

Рис. 6.25. Теплофизические свойства воды при нормальных условиях

Если округлить значения плотности, вязкости и удельной изобарной теплоемкости воды при нормальных условиях, то мы получим три единицы [6]. В статье, на которую сделана ссылка, раскрыты непривычные, но более правильные по мнению автора единицы вязкости и теплопроводности, показанные на рис. 6.25 и 6.26. А теплоемкость воды сам бог велел измерять калориями, а не джоулями. Калория – это энергия, которую нужно передать грамму воды, чтобы при нормальных условиях и постоянном давлении повысить её температуру на один кельвин. Все просто и понятно. А джоулями пусть измеряют механическую работу. Система единиц измерения, встроенная в пакет SMath, помимо прочего удобна и тем, что она позволяет вводить в расчет и выводить на печать значения физических величин с теми единицами, которые более привычны пользователю. В недрах же компьютера все будет храниться в единицах СИ. Подражая Козьме Пруткову, можно сказать так. Если вы видите в расчете запись 1 атм, то не верьте глазам своим. В этой клетке, пардон, в этом расчете хранится не атмосфера физическая, а паскаль. Это очень непривычно для тех, кто ведет расчеты вручную с помощью калькулятора или в средах «безразмерных» электронных таблиц и языков программирования [7].

Кстати, "легким движением руки" – кликаньем мышки по паре команд пакета SMath можно одним махом перевести русские названия единиц измерения, задействованные в наших расчетах, в английские (международные). Попутно можно заменить запятые, разделяющие целую и дробную части чисел, на точку (международный стандарт написания десятичных дробей). Это облегчает международный обмен информацией, в частности, обмен результатами расчетов.

Расчет на рис. 6.26 практически повторяет расчет на рис. 6.25. Только слово H2O (вода) заменено на слово Ethanol (этиловый спирт).

$$\text{CoolProp\_get\_fluid\_param\_string}(\text{"Ethanol"}; \text{"formula"}) = \text{"C\_\{2\}H\_\{6\}O\_\{1\}"}$$

$$\rho_{C2H5OH} := \text{CoolProp\_Props}(\text{"D"}; \text{"P"}; p; \text{"T"}; T; \text{"Ethanol"}) = 0{,}7894\ \frac{\text{г}}{\text{см}^3}$$

$$\text{CoolProp\_Props}(\text{"M"}; \text{"P"}; p; \text{"T"}; T; \text{"Ethanol"}) = 46{,}07\ \frac{\text{г}}{\text{моль}}$$

$$\text{CoolProp\_Props}(\text{"T"}; \text{"P"}; p; \text{"Q"}; 1; \text{"Ethanol"}) = 78{,}42\ ^{\circ}\text{C}$$

$$\text{CoolProp\_Props}(\text{"P"}; \text{"T"}; T; \text{"Q"}; 1; \text{"Ethanol"}) = 44{,}07\ \text{мм рт.ст.}$$

$$\text{CoolProp\_Props}(\text{"V"}; \text{"P"}; p; \text{"T"}; T; \text{"Ethanol"}) = 1{,}194\ \frac{\text{Н мм}}{\frac{\text{м}}{\text{с}}\ \text{м}^2}$$

$$\text{CoolProp\_Props}(\text{"Cpmass"}; \text{"P"}; p; \text{"T"}; T; \text{"Ethanol"}) = 0{,}5723\ \frac{\text{кал}}{\text{г К}}$$

$$\text{CoolProp\_Props}(\text{"conductivity"}; \text{"P"}; p; \text{"T"}; T; \text{"Ethanol"}) = 164{,}5\ \frac{\text{Вт мм}}{\text{м}^2\ \text{К}}$$

Рис. 6.26. Теплофизические свойства безводного этилового спирта

На рисунке 6.27 показан расчет параметров водки, полученной при смешении 40 мл этилового спирта и 60 мл воды[8]. После такой операции температура раствора немного повысится, но расчеты будут вестись для заданных 20 градусов по шкале Цельсия, так как все за короткое время вернется в стабильное температурное состояние.

Рассчитаны массы компонентов (m) готовой смеси (31,58 г[9] спирта и 59,89 г воды), а также массовые доли спирта (0,3452) и воды (0,6548), по которым формируется переменная *Водка*, участвующая в расчетах свойств этого спиртного напитка.

$$m_{C2H5OH} := 40\ \text{мл} \cdot \rho_{C2H5OH} = 31{,}58\ \text{г} \qquad m_{H2O} := 60\ \text{мл} \cdot \rho_{H2O} = 59{,}89\ \text{г}$$

$$m_{Водка} := m_{C2H5OH} + m_{H2O} = 91{,}47\ \text{г}$$

$$\frac{m_{C2H5OH}}{m_{Водка}} = 0{,}3452 \qquad \frac{m_{H2O}}{m_{Водка}} = 0{,}6548 \qquad 0{,}3456 + 0{,}6544 = 1$$

$$Водка := \text{"Ethanol[0.3456]\&Water[0.6544]"}$$

$$\rho_{Водка} := \text{CoolProp\_Props}(\text{"D"}; \text{"P"}; p; \text{"T"}; T; Водка) = 0{,}8508\ \frac{\text{г}}{\text{см}^3}$$

$$V_{Водка} := \frac{m_{Водка}}{\rho_{Водка}} = 107{,}5\ \text{мл}$$

Рис. 6.27. Расчет объема водки

Слегка переиначивая бессмертную поэму Венички Ерофеева "Москва-Петушки" можно сказать: "*… никто в России не знает, отчего умер Пушкин, а то, что при смешении 40 мл спирта и 60 мл воды выходит меньше, чем 100 мл водки – всякий знает*"[10]. А в нашем расчете получилось наоборот – водки получилось немного больше ста мililитров! Это, конечно, очень удачно с точки зрения какого-нибудь забулдыги, разбавляющего водой спиртовой настой боярышника. В пузырек этого настоя уж точно нужно сунуть ареометр или что-то ещё

[8] В перестроечные времена продавался знаменитый спирт Royal, который разбавляли водой. Как тут не вспомнить классическое булгаковское: "— Доктор Борменталь, умоляю вас, оставьте икру в покое! И если хотите послушаться доброго совета, налейте не английской, а обыкновенной русской водки.
– Новоблагословенная? - осведомился он.
– Бог с вами, голубчик, – отозвался хозяин. – Это спирт. Дарья Петровна сама отлично готовит водку."

[9] В ресторанах России, как правило, никогда не заказывают крепкие спиртные напитки в миллилитрах, всегда тут упоминают граммы, имея в виду не единицы массы, а единицы объема, вместимости: «Ещё по сто грамм на посошок и расходимся!».

[10] У Ерофеева немного иначе: "… никто в России не знает, отчего умер Пушкин, а как очищается политура – это всякий знает." Эта техническая жидкость благодаря высокому содержанию спирта может – после соответствующей очистки – употребляться в качестве суррогатного алкогольного напитка. Так что наше переиначивание Ерофеева не выбило нас из темы повествования.

и убедиться, что там этиловый, а не метиловый спирт. Пакет CoolProp Wrapper, кстати, показывает, что при наших 20 градусах по шкале Цельсия плотность этилового спирта составляет 0,7894 г/см^3, а метилового спирта 0,7913 г/см^3. Разница невелика и находится в пределах ошибки измерений. Неопределенность измерений, как стали сейчас говорить. Так что, лучше не пить того, что не предназначено для питья!

Но мы отвлеклись!

В чем тут дело!? Где запрятана ошибка! В вине должна храниться истина! Можно тут также вспомнить докторскую диссертацию Д.И. Менделеева "О соединении спирта с водой", которая обросла множеством легенд.

Если в интернете сделать запрос по ключу "Плотность водных растворов этанола", то будет выдано много информации, включая и таблицу, показанную на рис. 6.28. Кто-то любезно отсканировал её из книги Рабиновича В.А. и Хавина З.Я. "Краткий химический справочник" (3-е изд., СПб, Химия, 1991 г.) и разместил в интернете. Там же можно найти и другие такие таблицы из других бумажных справочников, сравнить их и убедиться, что они практически совпадают. Да, в интернете часто публикуется довольно сомнительная информация, которую всегда нужно проверять и перепроверять. Но и в бумажных справочниках, прошедших строгое рецензирование и редактирование, увы, встречаются ошибки и опечатки («очепятки»). Так, в таблице на рис. 6.28 указано, что плотность измеряется в г/см вместо правильной размерности г/см^3.

Таблица 6-6

Плотность водных растворов этилового спирта в зависимости от температуры, *г/см*

Весовой процент спирта	Температура, °C				Весовой процент спирта	Температура, °C			
	10	15	20	40		10	15	20	40
0	0,99973	0,99913	0,99823	0,99225	52	0,91723	0,91333	0,90936	0,89288
4	0,99218	0,99195	0,99103	0,98485	56	0,90831	0,90433	0,90031	0,88335
8	0,98660	0,98584	0,98478	0,97808	60	0,89927	0,89523	0,89113	0,87417
12	0,98145	0,98041	0,97910	0,97150	64	0,89006	0,88597	0,88183	0,86466
16	0,97692	0,97552	0,97387	0,96512	68	0,88074	0,87660	0,87241	0,85507
20	0,97252	0,97068	0,96864	0,95856	72	0,87127	0,86710	0,86287	0,84540
24	0,96787	0,96558	0,96312	0,95168	76	0,86168	9,85747	0,85322	0,83564
28	0,96268	0,95996	0,95710	0,94438	80	0,85197	0,84772	0,84344	0,82578
32	0,95655	0,95357	0,95038	0,93662	84	0,84203	0,83777	0,83348	0,81576
36	0,94986	0,94650	0,94306	0,92843	88	0,83181	0,82754	0,82323	0,80552
40	0,94238	0,93882	0,93518	0,91992	92	0,82114	0,81688	0,81257	0,79491
44	0,93433	0,93062	0,92685	0,91108	96	0,80991	0,80566	0,80138	0,78388
48	0,92593	0,92211	0,91823	0,90207	100	0,79784	0,79360	0,78934	0,77203

Рис. 6.28. Таблица плотности водных растворов этилового спирта

Есть в США такая организация – Институт стандартов и технологий НИСТ (www.nist.gov). Автор во время своего посещения этой научной организации в городе Боулдер штата Колорадо увидел там такую картину. В компьютерном классе попарно у мониторов сидели молодые люди и что-то там делали, шепчась и клацая клавишами клавиатуры. Автор удивился и спросил у американских коллег: "А что, у вас тоже не хватает вычислительной техники, и приходится у одного компьютера сажать двоих?". Но автору объяснили, что это студенты местного университета, подрабатывающие в НИСТ. Они вводят в компьютер информацию, подобную той, которая показана на рис. 6.28. Первый студент диктует числа из бумажной книги или научного журнала, а второй вводит их в компьютер. При этом первый студент дополнительно контролирует правильность ввода, попеременно поглядывая то в справочник, то на монитор компьютера. В НИСТ пытались было сканировать такие таблицы и оцифровывать их автоматически. Но плохое качество печатного оригинала и несовершенство ридеров[11] приводило к тому, что в итоговой оцифрованной таблице оказывалось много опечаток, которые приходилось выявлять и исправлять вручную. Проще и надежнее было сразу нанять студентов, дав им заодно возможность подработать. Обучение в вузах США платное – студентам приходится вертеться.

Автор на практических занятиях поручил своим студентам сделать такую работу на компьютере по технологии "щечка к щечке" – сравните рис. 6.28 и 6.29.

На рис. 6.29 показано содержимое матрицы M с "шапкой", хранящей температуру, и с "боковиком" (первым столбцом), хранящим весовой процент спирта. Это, конечно, не весовой, а массовый процент. Но в "старые добрые времена" эти две физические величины часто путали. Да и сейчас мы измеряем вес человека в килограммах, а не в более правильных ньютонах. Вес – это сила, а не масса! Но восклицание: "Во мне 1000 ньютонов, а не 102 килограмма!" помимо прочего попахивает манией величия.

[11] Программы, переводящие отсканированные тексты в цифровой формат.

$$M := \begin{bmatrix} \text{"n/t"} & 10 & 15 & 20 & 40 \\ 0 & 0.99973 & 0.99913 & 0.99823 & 0.99225 \\ 4 & 0.99218 & 0.99105 & 0.99103 & 0.98485 \\ 8 & 0.98660 & 0.98584 & 0.98478 & 0.97808 \\ 12 & 0.98145 & 0.98041 & 0.9791 & 0.97150 \\ 16 & 0.97692 & 0.97552 & 0.97387 & 0.96512 \\ 20 & 0.97252 & 0.97068 & 0.96864 & 0.95856 \\ 24 & 0.96787 & 0.96558 & 0.96312 & 0.95168 \\ 28 & 0.96268 & 0.95996 & 0.95710 & 0.94438 \\ 32 & 0.95655 & 0.95357 & 0.95038 & 0.93662 \\ 36 & 0.94986 & 0.9465 & 0.94306 & 0.92843 \\ 40 & 0.94238 & 0.93882 & 0.93518 & 0.91992 \\ 44 & 0.93433 & 0.93062 & 0.92685 & 0.91108 \\ 48 & 0.92593 & 0.92211 & 0.91823 & 0.90207 \\ 52 & 0.91723 & 0.91333 & 0.90936 & 0.89288 \\ 56 & 0.90831 & 0.90433 & 0.90031 & 0.88335 \\ 60 & 0.89927 & 0.89523 & 0.89113 & 0.87417 \\ 64 & 0.89006 & 0.88597 & 0.88183 & 0.86466 \\ 68 & 0.88074 & 0.8766 & 0.87241 & 0.85507 \\ 72 & 0.87127 & 0.8671 & 0.86287 & 0.84540 \\ 76 & 0.86168 & 0.85747 & 0.85322 & 0.83564 \\ 80 & 0.85197 & 0.84772 & 0.84344 & 0.82578 \\ 84 & 0.84203 & 0.83777 & 0.83348 & 0.81576 \\ 88 & 0.83181 & 0.82754 & 0.82323 & 0.80552 \\ 92 & 0.82114 & 0.81688 & 0.81257 & 0.79491 \\ 96 & 0.80991 & 0.80566 & 0.80138 & 0.78388 \\ 100 & 0.79784 & 0.7936 & 0.78934 & 0.77203 \end{bmatrix}$$

Рис. 6.29. Цифровой двойник бумажной таблицы, показанной на рис. 6.28

Матрицу M далее планировалось использовать в авторской программе [8] с двойной сплайн-интерполяцией, опирающейся на двойной цикл `for` (см. рис. 14.7, занятие 14 – Mathcad Prime [8]). Но на сайте пользователей SMath (https://en.smath.com/forum/yaf_postsm80962_lspline-and-linterp---where-is-an-error.aspx) подсказали, что есть более простое решение – см. рис. 6.30 с функцией `InterpBilinear`, которой нет в Mathcad, но которая есть в SMath.

$$\rho_{Vodka}(T;n) := \left| \begin{array}{l} x := \frac{T}{\mathrm{K}} - 273{,}15 \\ y := 100 \cdot n \\ X := \mathrm{submatrix}(M;1;1;2;\mathrm{cols}(M)) \\ Y := \mathrm{submatrix}(M;2;\mathrm{rows}(M);1;1) \\ Z := \mathrm{submatrix}(M;2;\mathrm{rows}(M);2;\mathrm{cols}(M)) \\ \mathrm{InterpBilinear}(X;Y;Z;x;y)\,\frac{\mathrm{г}}{\mathrm{см}^3} \end{array} \right.$$

Рис. 6.30. Программа двойной сплайн-интерполяции табличных данных

В программе на рисунке 6.30, во-первых, обрабатываются аргументы: температура переводится в градусы по шкале Цельсия ($x := T/\mathrm{K} - 273{,}15$), а массовая доля превращается в массовый процент ($y := 100\ n$) – все в соответствие с «боковиком» и «шапкой» исходной таблицы. На последней строке к рассчитанной безразмерной плотности добавляется нужная единица измерения. Это стандартные правила работы с эмпирическими формулами [9].

В программе на рис. 6.30 три раза вызывается встроенная в SMath функция `submatrix`, которая изымает из матрицы M «шапку» X с дискретными значениями температуры, «боковик» Y с дискретными значениями массовой доли и матрицу Z с дискретными значениями плотности (ядро матрицы M).

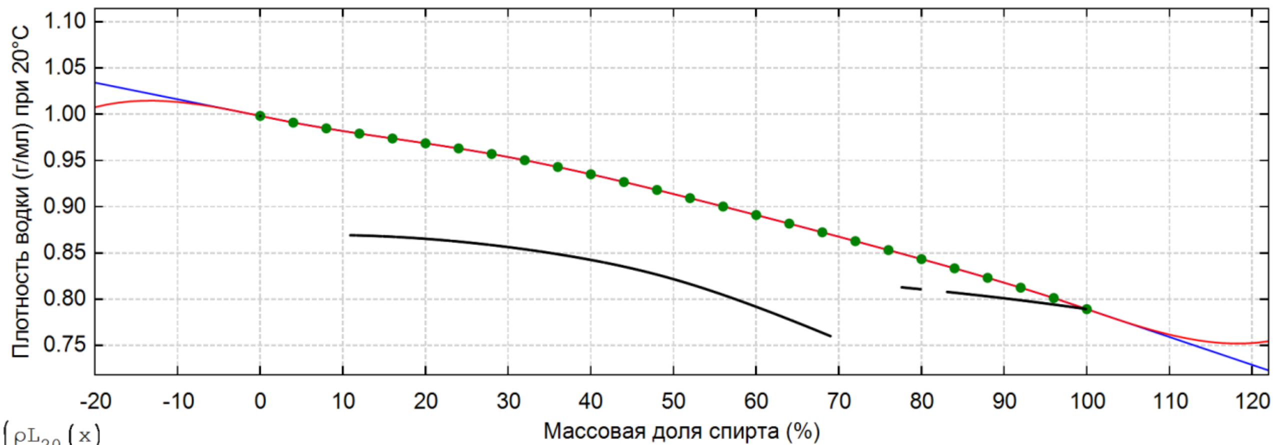
Функция `InterpBilinear` проводит кусочно-линейную интерполяцию, но не по одному (рис. 6.19), а по двум аргументам. Есть еще и интерполяция сплайнами (lspline), и мы её выше разобрали, нарисовав некое произведение искусства – см. рис. 6.23. На рисунках 6.31 и 6.32 показаны отличия в работе этих двух функций – синяя кривая – кусочно-линейная интерполяция, красная кривая – интерполяция сплайном. Зеленые точки – это дискретные данные, по которым велась интерполяция.

$vx := \mathrm{submatrix}(M, 2, \mathrm{rows}(M), 1, 1) \qquad vy := \mathrm{submatrix}(M, 2, \mathrm{rows}(M), 4, 4)$

$\rho S_{20}(x) := \mathrm{lspline}(vx, vy, x) \qquad \rho L_{20}(x) := \mathrm{linterp}(vx, vy, x)$

```
ρCP(x) := | Vodka := concat("Ethanol[", num2str(x/100, "n5"), "]&", "H2O[", num2str(1 − x/100, "n5"), "]")
          | try
          |   CoolProp_Props("D", "P", 1 atm, "T", 20 °C, Vodka) cm^3/g
          | on error
          |   0 cm^3/g
```

$xx := [0, 0.1 .. 100]$

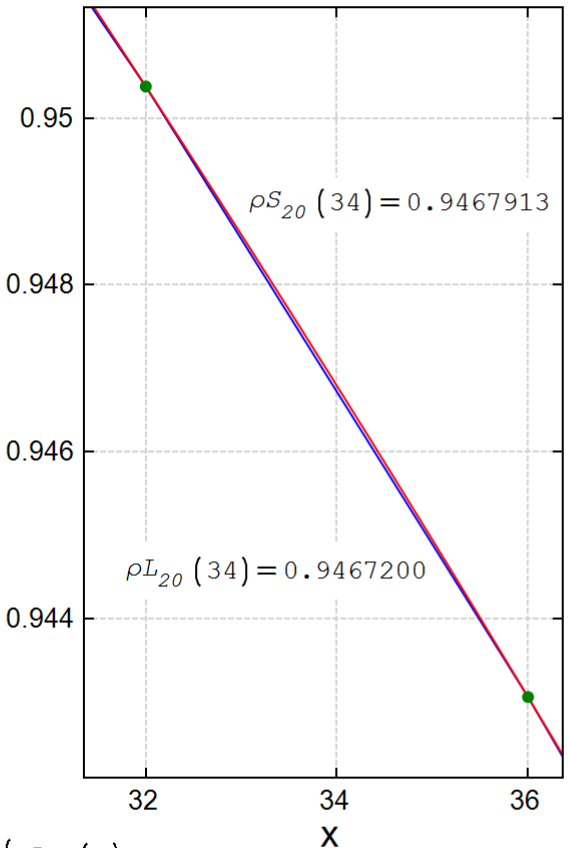


$$\begin{cases} \rho L_{20}(x) \\ \rho S_{20}(x) \\ \mathrm{augment}(vx, vy, ".", 5, "green") \\ \mathrm{augment}\left(xx, \overrightarrow{\rho_{CP}(xx)}, ".", 0.5, "black"\right) \end{cases}$$

Рис. 6.31. Изотерма 20°C: линейная интерполяция и интерполяция сплайнами

Внутри точек (интерполяция) синяя (функция ρL_{20}) и красная (ρS_{20}) кривые практически совпали. Существенно расходятся они в разные стороны вне зеленых точек (экстраполяция, которая в данной задаче не имеет физического смысла).

На рис. 6.32 показано увеличение масштаба кривых двух видов интерполяции вблизи двух точек, где есть небольшое расхождение. Но эти два вида интерполяции отличаются друг от друга не только количественно (см. рис. 6.32), но и качественно. Дело в том, что первая производная кусочно-линейной приближающей функции имеет точку разрыва первого рода в каждом узле интерполяции. Это усложняет работу с такой функцией при реализации различных численных методов. Такого недостатка лишена интерполяция сплайнами.



$$\begin{cases} \rho L_{20}(x) \\ \rho S_{20}(x) \\ \mathrm{augment}(vx, vy, ".", 5, "green") \end{cases}$$

Рис. 6.32. Кривые линейной интерполяции (синяя) и интерполяции сплайнами (красная)

На сайте https://community.ptc.com/t5/Mathcad/Spline/td-p/869944 описано, что означают буквы l, p и c перед словом spline в имени функции сплайн-интерполяции.

Дополнительно на графике рисунка 6.31 показана кривая, вернее отдельные обрывки кривой плотности водки, рассчитанной согласно пакету CoolProp Wrapper. Черная кривая – это ошибочные значения плотности, а разрывы в ней – это неудачные попытки (`try`) вызова функции ρ_{CP}, когда выдается не число, а сообщение об ошибке. Внутри крайней левой зеленой точки (чистая вода) находится черная точка - пакет CoolProp Wrapper здесь, как и в случае с чистым спиртом (крайняя правая зеленая точка), выдал правильное значение. В остальных точках было выдано неправильные значения.

Вывод. Инструменты "клевого" пакета CoolProp работы со смесью этанола с водой никуда не годятся. Скорее всего, такая же картина будет наблюдаться и при работе с другими смесями жидкостей. Например, воды и метилового спирта. И дело тут, по-видимому, в том, что эти инструменты предназначены для работы только с газами. А водка при нормальных условиях – это далеко не газ. Догадка авторов подтвердилась – на сайте пакета CoolProp Wrapper была «откопана информация мелким шрифтом» о том, что при работе со смесями наблюдаются большие проблемы. Здесь нужно было, конечно, заглушить такие расчеты. Но!

Частичный отказ от встроенной функции пакета CoolProp Wrapper и переход к "доморощенной" функции с именем ρ_{Vodka} дало правильный результат расчета объема водки (96,71 мл) – см. рис. 6.33. В свернутой области (линия с плюсиком слева) хранится функция, показанная на рис. 6.30, и опирающаяся на матрицу из рис. 6.29. Отказ частичный потому, что для определения плотности чистой воды и чистого этилового спирта мы использовали функцию пакета CoolProp Wrapper. В случае со смесью этого делать нельзя (см. рис. 6.31).

$$p := 1\ \text{atm} \qquad T := 20\ °\text{C}$$

$$\rho_{H2O} := \text{CoolProp\_Props}(\text{"D"}, \text{"P"}, p, \text{"T"}, T, \text{"Water"}) = 0.9982\ \frac{\text{g}}{\text{cm}^3}$$

$$\rho_{C2H5OH} := \text{CoolProp\_Props}(\text{"D"}, \text{"P"}, p, \text{"T"}, T, \text{"Ethanol"}) = 0.7894\ \frac{\text{g}}{\text{cm}^3}$$

$$V_{C2H5OH} := 40\ \text{mL} \qquad m_{C2H5OH} := V_{C2H5OH} \cdot \rho_{C2H5OH} = 31.58\ \text{g}$$

$$V_{H2O} := 100\ \text{mL} - V_{C2H5OH} = 60\ \text{mL} \qquad m_{H2O} := V_{H2O} \cdot \rho_{H2O} = 59.89\ \text{g}$$

$$m_{Vodka} := m_{C2H5OH} + m_{H2O} = 91.47\ \text{g}$$

$$n_{C2H5OH} := \frac{m_{C2H5OH}}{m_{Vodka}} = 0.3452 \qquad n_{H2O} := \frac{m_{H2O}}{m_{Vodka}} = 0.6548 \qquad n_{C2H5OH} + n_{H2O} = 1$$

⊞—A function of density of vodka depending on the T and mass fraction of ethyl alcohol

$$\rho_{Vodka}(T, n_{C2H5OH}) = 0.9458\ \frac{\text{g}}{\text{cm}^3} \qquad V_{Vodka} := \frac{m_{Vodka}}{\rho_{Vodka}(T, n_{C2H5OH})} = 96.71\ \text{mL}$$

Рис. 6.33. Правильный расчет объема водки

На рисунке 6.34 решается обратная задача: задана плотность водки (0,95 г/см^3) – найти массовую долю. Используется встроенная в SMath функция `solve`, возвращающая корень уравнения в заданном диапазоне неизвестного от нуля до единицы. Используется метод половинного деления.

$$n_{C2H5OH} := \mathrm{solve}\left(\rho_{Vodka}\left(T;\ n_{C2H5OH}\right) = 0,95\ \frac{\text{г}}{\text{см}^3};\ n_{C2H5OH};\ 0;\ 1\right) = 0,3221$$

Рис. 6.34. Расчет массовой доли спирта в разбавленной водке с плотностью 0,95 г/см^3

Несложно подсчитать, что водка с такой плотностью будет иметь крепость 37,5%. А водка, как говорил профессор Преображенский, должна иметь крепость 40 градусов. Хотя процент и градус тут не совпадают. Считается, что стоградусный этиловый спирт за счёт поглощения влаги из воздуха содержит примерно 96 процентов чистого спирта по объёму. Остальное – вода.

Задание читателям

1. Доработать расчёт на рис. 6.2 так, чтобы он учитывал веса элементов выборки, хранящихся в добавленной третьей строке матрицы `Data`.
2. Ввести в расчёт рост и размер обуви студентов (школьников) вашей учебной группы (класса) и провести регрессионный анализ.
3. Провести расчет, показанный на рис. 6.20-6.21, с использованием функции `cspline`.
4. На рисунке 6.21 решается система алгебраических уравнений с использованием функции `roots` для решения нелинейных систем. Но эта система линейная. Осуществите это решение с помощью встроенных инструментов SMath для решения систем линейных алгебраических уравнений.
5. Найдите в интернете данные по годам жизни великих и выдающихся физиков, писателей, художников и других деятелей науки и искусства и постройте для них соответствующие гистограммы (см. рис. 6.15).

Литература:

1. Очков В.Ф., Богомолова Е.П. Интерполяция, экстраполяция, аппроксимация или "Ложь, наглая ложь и статистика // Cloud of Science. Т. 2, № 1. 2015. С. 61-88 (http://twt.mpei.ac.ru/ochkov/CoS_2_1.pdf)

2. В.Ф. Очков, Ю.В. Чудова, Н.А. Очкова. Академическая шапочка математика, или Гибрид символа, числа и графика в задаче оптимизации // Математика в школе. № 4. 2020. С. (http://twt.mpei.ac.ru/ochkov/Math-School-01-20.pdf)
3. V. Ochkov, Y. Chudova. Academic Hats and Ice Cream: Two Optimization Problems // Journal of Humanistic Mathematics. Volume 12 Number 2 (July 2022 – http://twt.mpei.ac.ru/ochkov/Acad-Hat.pdf)
4. Лев Толстой и математика / В. Ф. Очков, Н. А. Очкова. – Москва: МПГУ, 2023. – 208 с.: ил. (http://twt.mpei.ac.ru/ochkov/Tolstoy-Math-3.pdf)
5. В. Ф. Очков, Н. А. Очкова. Проект памятника трем математикам или Матметрия // Cloud of Science. Том 4 № 4. 2017. С. 548-571 (http://twt.mpei.ac.ru/ochkov/MathMetria.pdf)
6. Очков В.Ф., Орлов К.А. Цифровой двойник воды // Энергия: экономика, техника, экология. № 10, 2021, С. 18-22 (http://www.twt.mpei.ac.ru/ochkov/Ochkov-3water.pdf)
7. Очков В.Ф. Физические и экономические величины в Mathcad и Maple. М.: Финансы и статистика, 2002 (http://www.twt.mpei.ac.ru/ochkov/Fis-Eco-Units-Mathcad-Maple.pdf)
8. Теплотехнические этюды с Excel, Mathcad и Интернет / Под общ. ред. В.Ф. Очкова. Издательство БХВ-Петербург. 2014. – 336 с. (последнее издание http://twt.mpei.ac.ru/ochkov/Therm-Studies.pdf)

Очков В.Ф., Орлов К.А. Единицы измерений в трех видах формул: в физических, эмпирических и… псевдоэмпирических // Мир измерений № 1-2, 2021 (http://twt.mpei.ac.ru/ochkov/Pseudo.pdf)