

ЗАНИМАТЕЛЬНАЯ СТАТИСТИКА

МАНГА

# РЕГРЕССИОННЫЙ АНАЛИЗ

Син Такахаши  
Иноуэ Ироха



**OHM**  
Ohmsha

**ОДЭКА**

**DMK**  
ИЗДАТЕЛЬСТВО

**Занимательная статистика**  
**РЕГРЕССИОННЫЙ АНАЛИЗ**

**Манга**

マンガでわかる

# 統計学

## [回帰分析編]

高橋 信 / 著  
井上 いろは / 作画  
トレンド・プロ / 制作



ОБРАЗОВАТЕЛЬНАЯ МАНГА

ЗАНИМАТЕЛЬНАЯ СТАТИСТИКА

# РЕГРЕССИОННЫЙ АНАЛИЗ

Син Такахаси  
Иноуэ Ироха



2-е издание, электронное

**ДМК**  
ПРЕСС  
ИЗДАТЕЛЬСТВО

**ОДЭКА**

Москва  
Додэка, ДМК Пресс, 2023

УДК 519.233.5

ББК 22.172

T15

**Такахаси, Син.**

T15 Занимательная статистика. Регрессионный анализ : манга / Такахаси Син (автор), Ироха Иноуэ (худож.) ; пер. с яп. А. Б. Клионского. — 2-е изд., эл. — 1 файл pdf : 224 с. — Москва : ДМК Пресс, Додэка-XXI, 2023. — (Образовательная манга). — Систем. требования: Adobe Reader XI либо Adobe Digital Editions 4.5 ; экран 10". — Текст : электронный.

ISBN 978-5-89818-409-4

Риса и Миу учатся в институте и подрабатывают в кафе Norns. Миу очень нравится один из посетителей кафе, но она не знает, как с ним познакомиться. Однажды он забывает на столе книгу, которая оказывается учебником по регрессионному анализу. Теперь есть повод заговорить с ним, но Миу почти не знакома с регрессионным анализом. И тогда Риса берётся помочь своей подруге разобраться в этой науке, осваивая которую Миу узнает, зачем и когда нужен регрессионный анализ, как его проводить, как оценивать его достоверность. В ходе обучения Миу познакомится с такими разновидностями регрессионного анализа, как множественная и логистическая регрессия и на примере кафе Norns научится практически их использовать.

Если у вас появилась необходимость на основе некоторых данных предсказать другие данные, например, зная расположение и площадь нового магазина предсказать, какая у него будет выручка, то эта книга для вас.

УДК 519.233.5

ББК 22.172

**Электронное издание на основе печатного издания:** Занимательная статистика. Регрессионный анализ : манга / Такахаси Син (автор), Ироха Иноуэ (худож.) ; пер. с яп. А. Б. Клионского. — Москва : ДМК Пресс, Додэка-XXI, 2014. — 214 с. — (Образовательная манга). — ISBN 978-5-97060-115-0. — Текст : непосредственный.

Издательство выражает благодарность *В. О. Панфилову*

Все права защищены. Никакая часть этого издания не может быть воспроизведена в любой форме или любыми средствами, электронными или механическими, включая фотографирование, ксерокопирование или иные средства копирования или сохранения информации, без письменного разрешения издательства.

В соответствии со ст. 1299 и 1301 ГК РФ при устранении ограничений, установленных техническими средствами защиты авторских прав, правообладатель вправе требовать от нарушителя возмещения убытков или выплаты компенсации.

ISBN 978-5-89818-409-4

© 2005 by Shin Takahashi and Trend-Pro Co., Ltd.

© Перевод, Издательский дом «Додэка-XXI», 2013

© Оформление, издание, ДМК Пресс, 2014

# ПРЕДИСЛОВИЕ

Данная книга посвящена описанию регрессионного анализа, множественной регрессии и логистической регрессии.

Регрессионный анализ и множественная регрессия — это методы анализа, которые позволяют предсказывать числовые значения. Они позволяют предсказать, например, число заказов чая со льдом на основе максимальной температуры воздуха или месячную выручку нового магазина на основе его площади и расстояния до ближайшей ж/д станции.

Логистическая регрессия — это метод анализа, предназначенный для прогноза вероятности. Он позволяет предсказать, например, вероятность заболевания раком на основе числа выкуренных сигарет или выпитого спиртного.

Для лучшего понимания книги желательно, чтобы читатели уже ознакомились с предыдущим произведением данной серии — «Занимательная статистика. Манга»\*, или же их знания в этой области не уступали разъяснённым в предыдущей манге.

Книга будет полезна людям, которым нужно прогнозировать числовые значения или вероятности. Скажу прямо: математический уровень данной книги выше, чем у «Занимательная статистика. Манга».

Книга имеет следующую структуру:

- Глава 1. Базовые знания;
- Глава 2. Регрессионный анализ;
- Глава 3. Множественная регрессия;
- Глава 4. Логистическая регрессия.

Каждая из глав, в свою очередь, состоит из собственно манги и текстового дополнения.

В первой главе рассказывается про то, что многие читатели, вероятно, изучали в курсе старшей школы, например, про производные и матрицы. Эти знания будут необходимы для понимания глав, начиная со второй. Но читателю не нужно бояться, что без понимания первой главы он не сможет прочесть последующие главы. Данную книгу можно читать легко, думая при этом: «Вроде бы, смысл логарифмов такой», или «Производная, вроде бы, находится вот так. Ура, вспомнил!» Однако читателям, которые скажут: «Я слишком забывчив, поэтому знания не всплывают в моей голове при чтении», или «У меня гуманитарное образование, и мы это вообще не изучали», нужно будет приложить некоторые усилия, чтобы понять содержание первой главы. В противном случае, чтение начиная со второй главы будет для них мучением.

---

\* Син Такахаси. Занимательная статистика. Манга. М.: Додэка. 2010. 224 с.

Читателю, сведущему в математике, будет полезно внимательно изучить подробные описания процесса расчётов, которые есть в данной книге. Тому, кто не особо силён в математике, можно просто бегло прочитать их. Другими словами, читателю вполне достаточно думать: «Да, смысл этого метода мне не понятен, вычисления выглядят сложно, но, тем не менее, это вроде бы позволяет найти решение». Не нужно пытаться понять всё сразу. Давайте заниматься спокойно, без спешки. Однако автор просит читателя обязательно пробежаться взглядом даже по непонятным объяснениям!

В связи с округлением, некоторые результаты вычислений, приведённые в книге, не совсем совпадут с результатами, которые читатели получают самостоятельно.

В заключение я хотел бы здесь выразить благодарность коллективу Отдела работок издательства Ohmsha, которые дали мне шанс взяться за перо. Я благодарю также сотрудников компании Trend-Pro, благодаря усилиям которых моя рукопись смогла стать мангой, господина re\_akino, отвечавшего за написание сценария, а также господина Иноуэ Ироха, художника этой манги. Кроме того, я глубоко благодарен господину Сакаори Фумитакэ, преподавателю факультета социологии университета Риккё, от которого я на этот раз получил даже больше ценных советов, чем при написании предыдущей книги.

Син Такахаси

# СОДЕРЖАНИЕ

Пролог.	
ДОБРО ПОЖАЛОВАТЬ В NORNIS! .....	1
Глава 1.	
БАЗОВЫЕ ЗНАНИЯ .....	11
1. Правила записи .....	12
2. Обратные функции .....	14
3. Показательные функции и функция натурального логарифма .....	19
4. Свойства показательных и логарифмических функций .....	20
5. Производная .....	24
6. Матрицы .....	37
7. Количественные и качественные данные .....	46
8. Сумма квадратов отклонений, рассеяние, стандартное отклонение ..	48
9. Функция плотности вероятности .....	50
Глава 2.	
РЕГРЕССИОННЫЙ АНАЛИЗ .....	55
1. Что такое регрессионный анализ .....	56
2. Конкретный пример регрессионного анализа .....	62
3. Замечание по поводу проведения регрессионного анализа .....	94
4. Нормированный остаток .....	95
5. Интерполяция и экстраполяция .....	96
6. Автокорреляция .....	97
7. Нелинейные уравнения регрессии .....	98
Глава 3.	
МНОЖЕСТВЕННАЯ РЕГРЕССИЯ .....	101
1. Что такое множественная регрессия .....	102
2. Конкретный пример множественной регрессии .....	106
3. Замечание по поводу проведения множественной регрессии .....	136
4. Нормированный остаток .....	137
5. Обобщённое расстояние Махаланобиса, доверительный интервал и прогнозируемый интервал .....	138
6. Множественная регрессия при наличии неизмеряемых данных среди объясняющих переменных .....	141

7. Мультиколлинеарность .....	145
8. Степень влияния каждой из объясняющих переменных на отклик и множественная регрессия .....	146

#### Глава 4.

### ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ..... 149

1. Что такое логистическая регрессия.....	150
2. Метод максимального правдоподобия .....	156
3. Трактовка отклика .....	160
4. Конкретный пример логистической регрессии.....	164
5. Замечание по поводу проведения логистической регрессии .....	186
6. Отношение перевесов .....	186
7. Пузырьковая диаграмма .....	192

#### Приложение.

### ПРОБУЕМ ВЫЧИСЛЯТЬ В EXCEL! .....

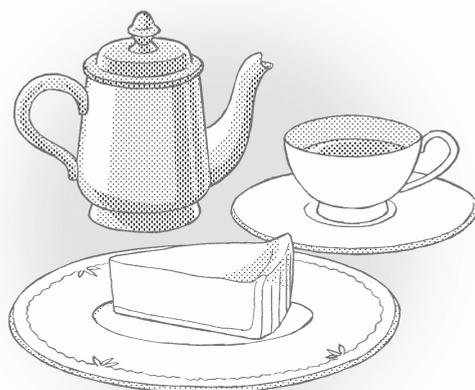
1. Основание натурального логарифма.....	194
2. Значение показательной функции.....	195
3. Значение функции натурального логарифма .....	196
4. Произведение матриц.....	197
5. Обратная матрица .....	199
6. Нахождение на оси x графика распределения хи-квадрат значения, соответствующего заданной вероятности.....	200
7. Вероятность распределения хи-квадрат.....	201
8. Нахождение на оси x графика F-распределения значения, соответствующего заданной вероятности .....	202
9. Вероятность F-распределения .....	204
10. (Частные) коэффициенты (множественной) регрессии .....	205
11. Коэффициенты уравнения логистической регрессии .....	208

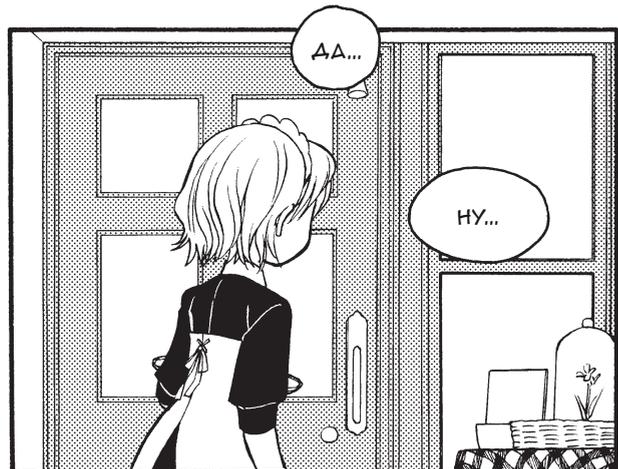
### Предметный указатель .....

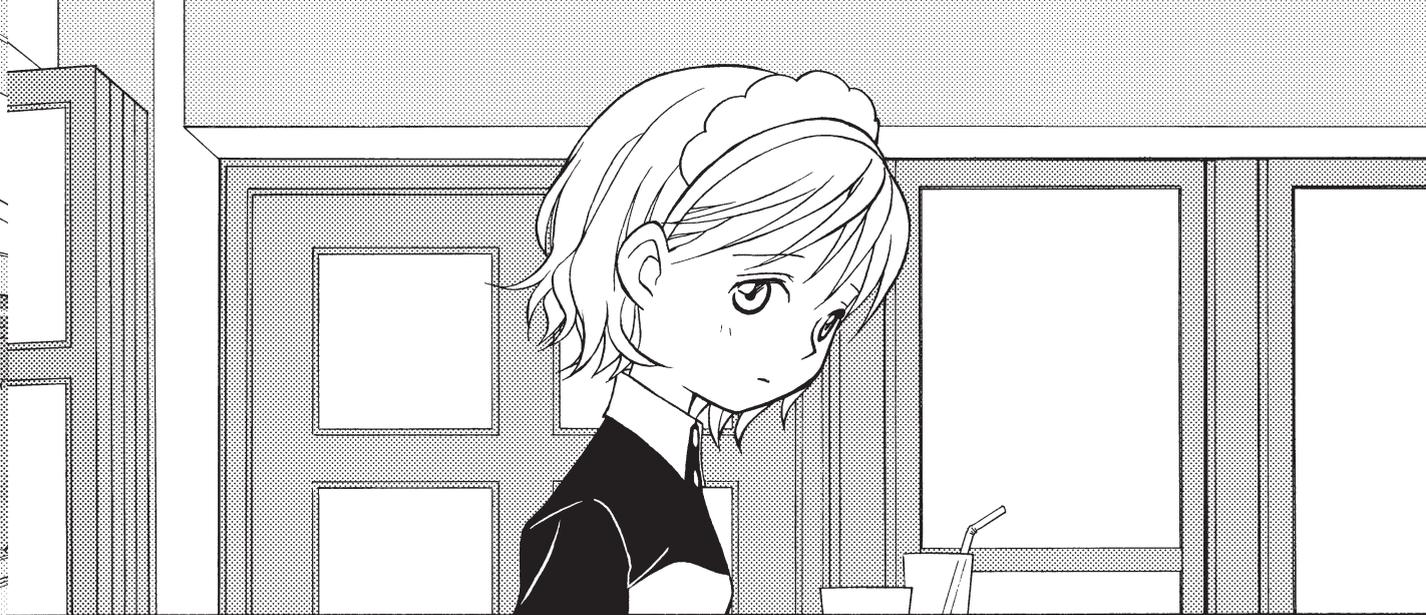
212

ПРОЛОГ

# ДОБРО ПОЖАЛОВАТЬ В NORNS!

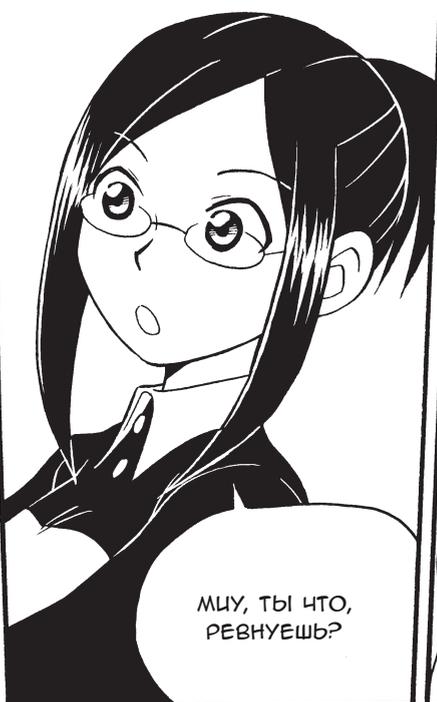








РИСА, КАК ТЫ  
С НИМ МИЛО  
РАЗГОВАРИВАЛА.

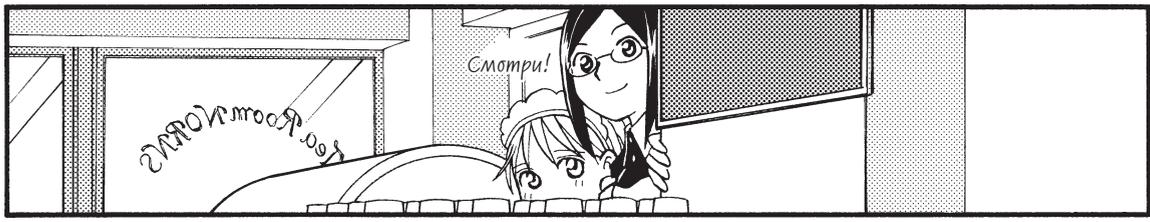


МУУ, ТЫ ЧТО,  
РЕВНУЕШЬ?



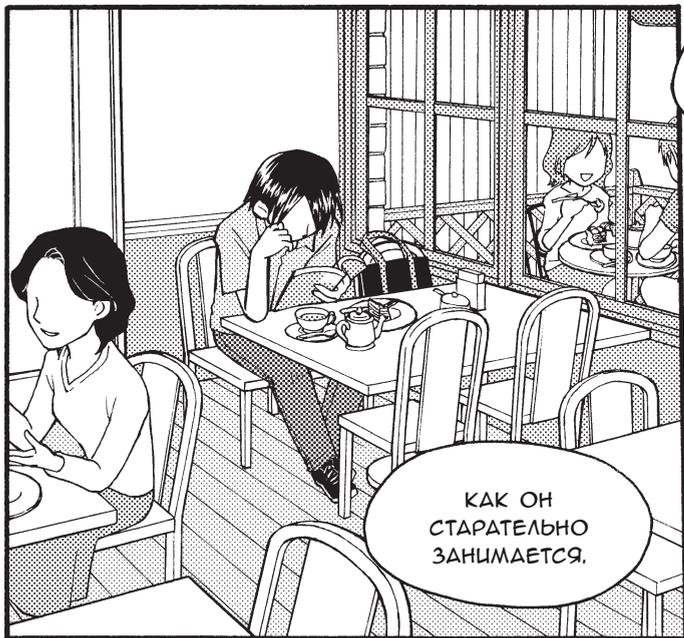
НЕТ, НИЧЕГО  
ПОДОБНОГО!

Да?  
Да?



Смотри!

НORNS

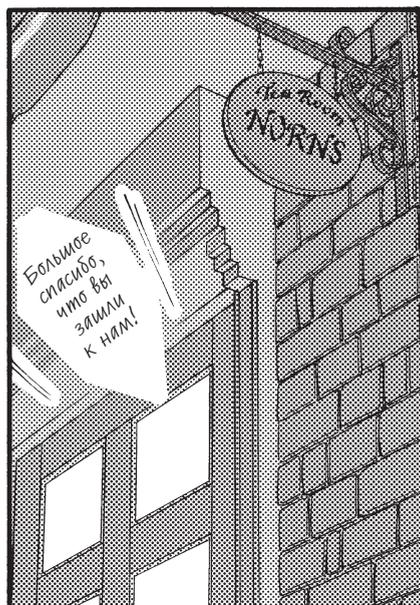


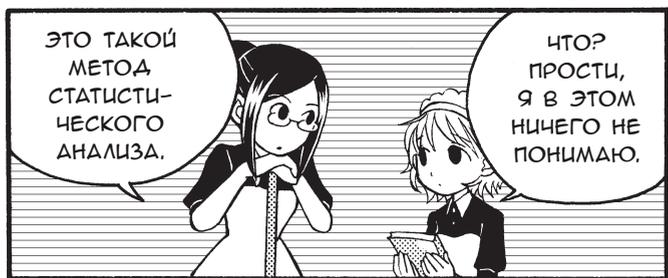
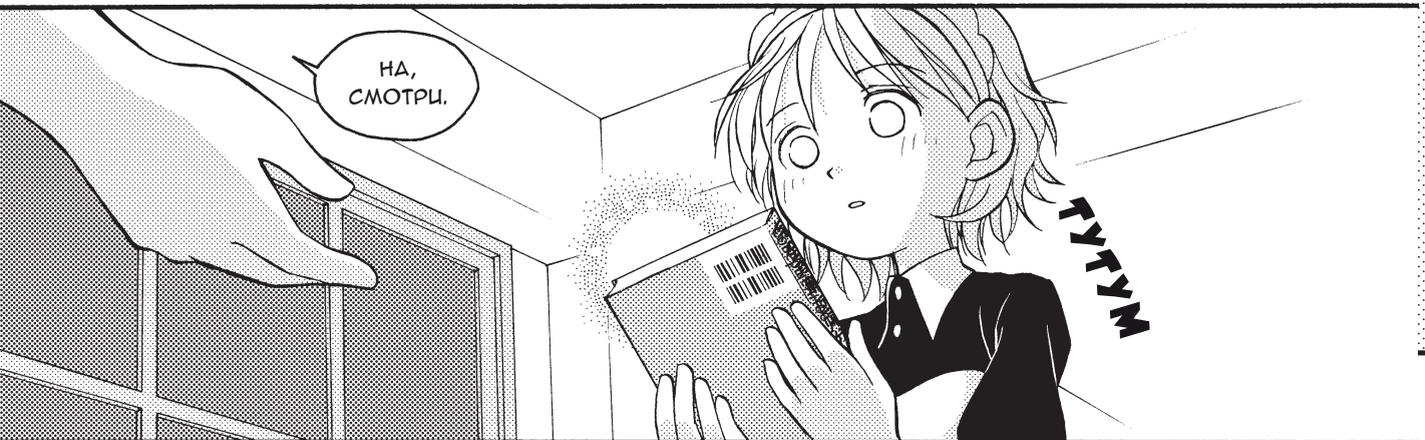
КАК ОН  
СТАРАТЕЛЬНО  
ЗАНИМАЕТСЯ.



ВСЁ ВРЕМЯ ЧИТАЕТ  
СЛОЖНЫЕ КНИГИ  
ПО МАТЕМАТИКЕ.

ПОСТОЙ!,  
ВЕДЬ МЫ ТОЖЕ  
ВРОДЕ С ЭКОНО-  
МИЧЕСКОГО ФА-  
КУЛЬТЕТА.





Сегодня воздух прогреется до 31°C

5°C

ПРЕДСТАВЬ, ЧТО МЫ В NORN'S КАЖДЫЙ ДЕНЬ ЗАПИСЫВАЕМ МАКСИМАЛЬНУЮ ТЕМПЕРАТУРУ ВОЗДУХА И ЧИСЛО ЗАКАЗОВ ЧАЯ СО ЛЬДОМ.

Так, запишем

8/22	29	77
8/23	28	62
8/24	34	93
8/25	31	84
8/26	25	59
8/27	29	64
8/28	32	80
8/29	31	

Регрессионный анализ

Сегодня воздух прогреется до 27°C

Бумажка!

Сегодня будет 65 стаканов чая со льдом!

Чай со льдом!

Чай со льдом!

3+3+2+1+1+4+3  
+2+1+1+1+3+2+3  
+2+3+...

РЕГРЕССИОННЫЙ АНАЛИЗ ПО МАКСИМАЛЬНОЙ ТЕМПЕРАТУРЕ ПРЕДСКАЗАТЬ ЧИСЛО ЗАКАЗОВ ЧАЯ СО ЛЬДОМ!

ПРАВДА? ВОТ ЗАОРОВО!

КРОМЕ ТОГО, СУЩЕСТВУЕТ ЕЩЁ ПОХОЖИЙ МЕТОД МНОЖЕСТВЕННОЙ РЕГРЕССИИ.

МНОЖЕСТВЕННАЯ?

Игра слов...

Игра слов...

ТОЛЬКО ЧТО МЫ ПРЕДСКАЗАЛИ ЧИСЛО ЗАКАЗОВ ПО ОДНОМУ ФАКТОРУ — "МАКСИМАЛЬНОЙ ТЕМПЕРАТУРЕ".

РЕГРЕССИОННЫЙ АНАЛИЗ — ЭТО ПРОГНОЗ ПО ОДНОМУ ФАКТОРУ, А МНОЖЕСТВЕННАЯ РЕГРЕССИЯ — ЭТО ПРОГНОЗ ПО МНОЖЕСТВУ ФАКТОРОВ.

Фактор

Прогноз

Факторы

Прогноз

Регрессионный анализ

Множественный регрессионный анализ

$x_1$

$x_2$   $x_3$

$y$

ДАВАЙ Я ПРИВЕДУ ПРИМЕР. ПУСТЬ ДИРЕКТОР ОДНОЙ СЕТИ РЕСТОРАНОВ ИМЕЕТ СЛЕДУЮЩИЕ ДАННЫЕ О КАЖДОМ ИЗ СВОИХ ЗАВЕДАНИЙ:

- УДАЛЕНИЕ ОТ РЕСТОРАНА КОНКУРИРУЮЩЕЙ СЕТИ
- ЧИСЛО ЖИЛЫХ ДОМОВ В РАДИУСЕ 500 МЕТРОВ
- РАСХОДЫ НА РЕКЛАМУ

	Удаление от конкурента, м	Число домов в радиусе 500 м	Расходы на рекламу, иены	Выручка, иены
A 	○○○	○○○	○○○	○○○
B 	△△△	△△△	△△△	△△△
C 	□□□	□□□	□□□	□□□
	⋮	⋮	⋮	⋮



Наша сеть маловата...

ЕСЛИ ОН ЗАХОЧЕТ ОТКРЫТЬ НОВЫЙ РЕСТОРАН,...



Вариант размещения

Так какой же вариант лучший?

...ТО МНОЖЕСТВЕННАЯ РЕГРЕССИЯ НА ОСНОВЕ:

- УДАЛЕНИЯ ОТ РЕСТОРАНА КОНКУРЕНТА
- ЧИСЛА ЖИЛЫХ ДОМОВ В РАДИУСЕ 500 МЕТРОВ
- РАЗМЕРА РАСХОДОВ НА РЕКЛАМУ

ПОЗВОЛИТ ЕМУ ПРЕДСКАЗАТЬ ВЫРУЧКУ НОВОГО МАГАЗИНА!



Ага! Лучшее место открыть его здесь!

ДА, УДОБНАЯ ШТУКА!

КРОМЕ ЭТОГО, ЕЩЁ **ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ.**

АА, КАК МНОГО ВСЕГО ЕСТЬ!



СПРАВЛЮСЬ ЛИ Я...?



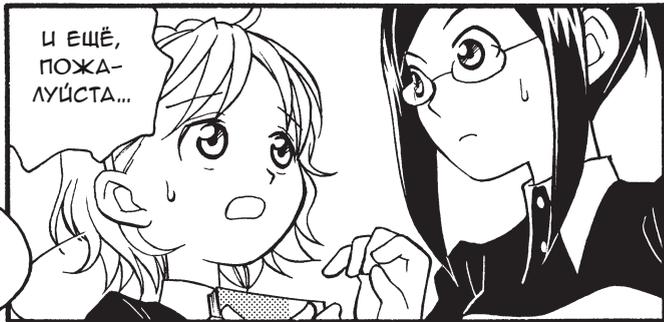
МОЖЕТ БЫТЬ, ЭТА КНИГА...

...ПОМОЖЕТ МНЕ СЕБЕЛИЗИТЬСЯ С НИМ!





Я ХОТЕЛА БЫ  
ВЗЯТЬ ЭТУ КНИГУ  
НА ХРАНЕНИЕ.



И ЕЩЁ,  
ПОЖА-  
ЛУЙСТА...

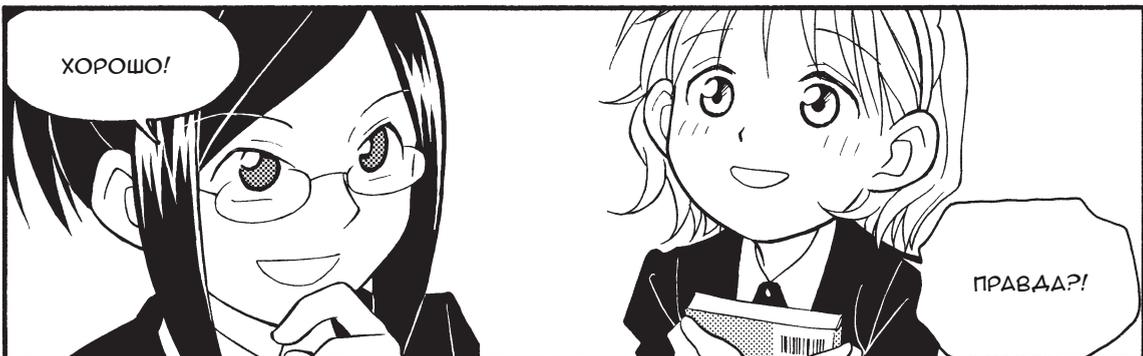


...НАУЧИ МЕНЯ  
РЕГРЕССИОННОМУ  
АНАЛИЗУ!



ХМ...

УМОЛЯЮ  
ТЕБЯ!



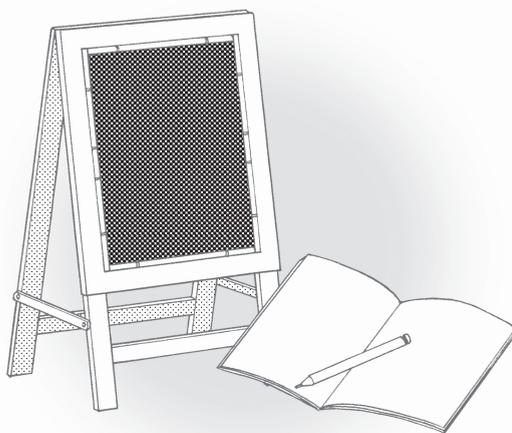
ХОРОШО!

ПРАВДА?!

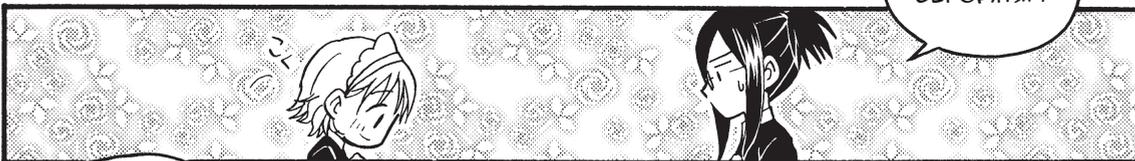


ГЛАВА 1

# БАЗОВЫЕ ЗНАНИЯ



# 1. ПРАВИЛА ЗАПИСИ





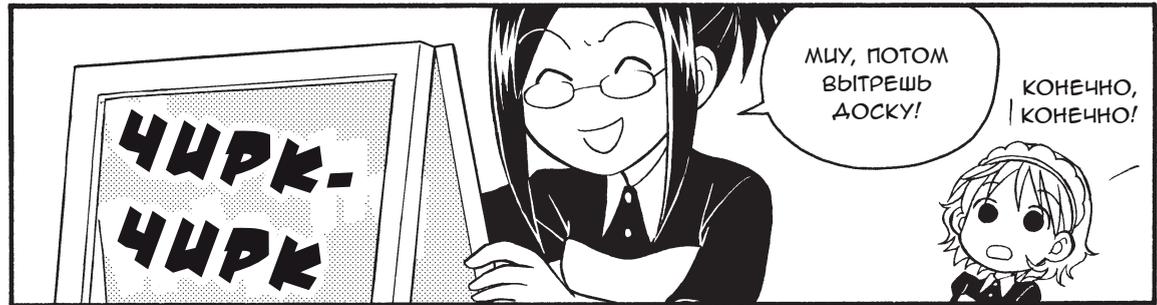
НУ, ХОРОШО,  
ПРИСТУПИМ!  
СЕГОДНЯ Я ДАМ ТЕБЕ  
БАЗОВЫЕ ЗНАНИЯ,  
БЕЗ КОТОРЫХ  
РАЗОБРАТЬСЯ  
В РЕГРЕССИОННОМ  
АНАЛИЗЕ БУДЕТ  
СЛОЖНОВАТО.

ХОРОШО, Я  
ГОТОВА!



ДЛЯ НАЧЛА —  
О ПРАВИЛАХ  
ЗАПИСИ, ПРИНЯТЫХ  
В МАТЕМАТИКЕ...

Сегодня  
в Меню:  
ДОСКА ДЛЯ  
МЕНЮ...



**ЧИРК-  
ЧИРК**

МЦУ, ПОТОМ  
ВЫТРЕШЬ  
ДОСКУ!

КОНЕЧНО,  
КОНЕЧНО!



**Правила записи**

⋮	⋮	⋮
$x \times x \times x$	$=$	$x^3$
$x \times x$	$=$	$x^2$
$x$	$=$	$x^1$
$1$	$=$	$x^0$
$\frac{1}{x}$	$=$	$x^{-1}$
$\frac{1}{x^2}$	$=$	$x^{-2}$
$\frac{1}{x^3}$	$=$	$x^{-3}$
⋮	⋮	⋮

ТАК КАК ЭТО —  
ПРАВИЛА, ТЫ ДОЛЖНА  
ПРИНЯТЬ ИХ БЕЗ  
МАЛЕЙШИХ СОМНЕНИЙ.

ЯСНО!

