Анализатор регрессии и корреляции

Математическая документация

Русский

Вольфганг Бёме
wolfgang.boehme@economic-modeling.de
www.economic-modeling.de

6 ноября 2026 г.

Документация на этом сайте разработана при поддержке ИИ Claude (Anthropic).

Содержание

1 Описательная статистика **3**
1.1 Меры положения . . . 3
1.1.1 Среднее арифметическое . . . 3
1.1.2 Медиана . . . 3
1.1.3 Квантили . . . 3
1.2 Меры рассеяния . . . 3
1.2.1 Дисперсия . . . 3
1.2.2 Стандартное отклонение . . . 4
1.2.3 Межквартильный размах (IQR) . . . 4
1.3 Параметры формы . . . 4
1.3.1 Асимметрия (Skewness) . . . 4
1.3.2 Эксцесс (Kurtosis) . . . 4

2 Корреляционный анализ **4**
2.1 Корреляция Пирсона . . . 5
2.2 Ранговая корреляция Спирмена . . . 5

3 Простая линейная регрессия **5**
3.1 Модель . . . 5
3.2 Оценка параметров (МНК) . . . 5
3.3 Коэффициент детерминации (R^2) . . . 6
3.4 Меры ошибки . . . 6
3.4.1 Среднеквадратичная ошибка (RMSE) . . . 6
3.4.2 Средняя абсолютная ошибка (MAE) . . . 6

4 Полиномиальная регрессия **6**
4.1 Модель степени p . . . 6
4.2 Матричная запись . . . 6
4.3 Нормальные уравнения . . . 6

5 Множественная линейная регрессия **7**
5.1 Модель . . . 7
5.2 Скорректированный R^2 . . . 7
5.3 Мультиколлинеарность: фактор инфляции дисперсии (VIF) 7

6 Нелинейные модели регрессии **7**
6.1 Экспоненциальная модель . . . 7
6.2 Степенная модель . . . 7
6.3 Логарифмическая модель . . . 8
6.4 Логистическая функция . . . 8
6.5 Функция Гомпертца . . . 8
6.6 Функция Гаусса . . . 8

7 Анализ остатков **8**
7.1 Остатки . . . 8
7.2 Стандартизированные остатки . . . 8
7.3 Рычаг (Leverage) . . . 9

7.4 Расстояние Кука 9
7.5 Q-Q график 9

8 Выбор модели **9**
8.1 Информационный критерий Акаике (AIC) 9

9 Пример 1: функция потребления США 1960–2008 **10**
9.1 Набор данных 10
9.2 Описательная статистика 10
9.3 Корреляционный анализ 11
9.4 Простая линейная регрессия: кейнсианская функция потребления 11
9.5 Анализ остатков 12
9.6 Множественная регрессия 12
9.7 Сравнение моделей: рост во времени 12

10 Пример 2: кривая Энгеля **13**
10.1 Степенная модель и эластичность по доходу 13
10.2 Множественная регрессия с размером домохозяйства . . . 13

11 Пример 3: производственная функция Кобба–Дугласа **13**
11.1 Линеаризация и множественная регрессия 14
11.2 Интенсивная форма 14

12 Итог: сравнение моделей **14**

13 Литература **15**

14 Приложение: численные методы **15**
14.1 Метод Гаусса 15
14.2 Аппроксимация обратной функции ошибок 15

Аннотация

Данная документация описывает математические основы и формулы анализатора регрессии и корреляции. Формально определены все реализованные статистические методы, включая описательную статистику, корреляционный анализ, линейные и нелинейные модели регрессии, а также диагностические процедуры. Три экономических примера — функция потребления (реальные данные), кривая Энгеля и производственная функция Кобба–Дугласа — иллюстрируют применение методов; они включены в инструмент как примеры наборов данных.

1. Описательная статистика

Пусть $X = \{x_1, x_2, \ldots, x_n\}$ – выборка из n наблюдений.

1.1. Меры положения

1.1.1. Среднее арифметическое

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1}$$

1.1.2. Медиана

Медиана определяется как:

$$\text{Медиана}(X) = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{если } n \text{ нечётно} \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & \text{если } n \text{ чётно} \end{cases} \tag{2}$$

где $x_{(i)}$ обозначает i-ю порядковую статистику (отсортированные значения).

1.1.3. Квантили

p-квантиль ($0 \leq p \leq 1$) определяется линейной интерполяцией:

$$Q_p = x_{(k)} + (x_{(k+1)} - x_{(k)}) \cdot r \tag{3}$$

где $h = p(n-1)$, $k = \lfloor h \rfloor$ и $r = h - k$.

1.2. Меры рассеяния

1.2.1. Дисперсия

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{4}$$

1.2.2. Стандартное отклонение

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{5}$$

1.2.3. Межквартильный размах (IQR)

$$\text{IQR} = Q_{0.75} - Q_{0.25} \tag{6}$$

1.3. Параметры формы

1.3.1. Асимметрия (Skewness)

Асимметрия измеряет несимметричность распределения:

$$\gamma_1 = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3 \tag{7}$$

Интерпретация:

- $\gamma_1 \approx 0$: симметричное распределение
- $\gamma_1 > 0$: правосторонняя асимметрия (правый хвост длиннее)
- $\gamma_1 < 0$: левосторонняя асимметрия (левый хвост длиннее)

1.3.2. Эксцесс (Kurtosis)

Эксцесс измеряет «тяжесть хвостов»:

$$\gamma_2 = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{8}$$

Интерпретация (избыточный эксцесс):

- $\gamma_2 \approx 0$: мезокуртическое (как нормальное распределение)
- $\gamma_2 > 0$: лептокуртическое (тяжёлые хвосты, острая вершина)
- $\gamma_2 < 0$: платикуртическое (лёгкие хвосты, плоская вершина)

2. Корреляционный анализ

Пусть $X = \{x_1, \ldots, x_n\}$ и $Y = \{y_1, \ldots, y_n\}$ – две переменные.

2.1. Корреляция Пирсона

Коэффициент корреляции Пирсона измеряет линейную зависимость:

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{9}$$

Эквивалентная формула:

$$r_{XY} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2} \cdot \sqrt{n\sum y_i^2 - (\sum y_i)^2}} \tag{10}$$

Выполняется: $-1 \leq r_{XY} \leq 1$

2.2. Ранговая корреляция Спирмена

Для рангов ранг(x_i) и ранг(y_i):

$$\rho_{XY} = r_{\text{ранг}(X),\text{ранг}(Y)} \tag{11}$$

Это коэффициент корреляции Пирсона для рангов, устойчивый к выбросам.

3. Простая линейная регрессия

3.1. Модель

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n \tag{12}$$

где $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ н.о.р.

3.2. Оценка параметров (МНК)

Минимизация суммы квадратов ошибок:

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \tag{13}$$

Оценки:

$$\hat{\beta}_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \tag{14}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{15}$$

3.3. Коэффициент детерминации (R^2)

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{16}$$

где $\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2$ – общая сумма квадратов.

Интерпретация: R^2 показывает долю объяснённой дисперсии ($0 \leq R^2 \leq 1$).

3.4. Меры ошибки

3.4.1. Среднеквадратичная ошибка (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{17}$$

3.4.2. Средняя абсолютная ошибка (MAE)

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{18}$$

4. Полиномиальная регрессия

4.1. Модель степени p

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p + \varepsilon_i \tag{19}$$

4.2. Матричная запись

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{20}$$

где

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & \cdots & x_2^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^p \end{pmatrix} \tag{21}$$

4.3. Нормальные уравнения

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{22}$$

Решение находится методом гауссова исключения.

5. Множественная линейная регрессия

5.1. Модель

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \tag{23}$$

5.2. Скорректированный R²

Скорректированный R^2 учитывает число параметров:

$$R^2_{\text{adj}} = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1} \tag{24}$$

где p – число предикторов.

5.3. Мультиколлинеарность: фактор инфляции дисперсии (VIF)

Для каждого предиктора x_j:

$$\text{VIF}_j = \frac{1}{1 - R_j^2} \tag{25}$$

где R_j^2 – коэффициент детерминации регрессии x_j на все остальные предикторы.

Интерпретация:

- $\text{VIF}_j < 5$: мультиколлинеарность не вызывает беспокойства
- $5 \leq \text{VIF}_j \leq 10$: умеренная мультиколлинеарность
- $\text{VIF}_j > 10$: высокая мультиколлинеарность (проблематично)

6. Нелинейные модели регрессии

6.1. Экспоненциальная модель

$$y = a \cdot e^{bx} \tag{26}$$

Линеаризация: $\ln(y) = \ln(a) + bx$

6.2. Степенная модель

$$y = a \cdot x^b \tag{27}$$

Линеаризация: $\ln(y) = \ln(a) + b \ln(x)$

6.3. Логарифмическая модель

$$y = a + b\ln(x) \tag{28}$$

6.4. Логистическая функция

$$y = y_{\min} + \frac{L}{1 + e^{-k(x-x_0)}} \tag{29}$$

где:

- L: ёмкость (разница между максимумом и минимумом)
- k: скорость роста
- x_0: точка перегиба

6.5. Функция Гомпертца

$$y = a \cdot e^{-b \cdot e^{-cx}} \tag{30}$$

Асимметричная кривая роста, часто применяется в биологии.

6.6. Функция Гаусса

$$y = a \cdot e^{-\frac{(x-b)^2}{2c^2}} \tag{31}$$

Симметричная колоколообразная кривая с параметрами:

- a: амплитуда (максимум)
- b: центр (среднее значение)
- c: ширина (стандартное отклонение)

7. Анализ остатков

7.1. Остатки

$$e_i = y_i - \hat{y}_i, \quad i = 1, \ldots, n \tag{32}$$

7.2. Стандартизированные остатки

$$r_i = \frac{e_i}{s_e} \tag{33}$$

где $s_e = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$ – стандартное отклонение остатков.

7.3. Рычаг (Leverage)

Для простой линейной регрессии:

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \tag{34}$$

Выполняется: $0 \leq h_i \leq 1$ и $\sum_{i=1}^{n} h_i = 2$ (простая регрессия).

Пороговое значение: $h_i > 2p/n$ указывает на высокий рычаг (где p – число параметров).

7.4. Расстояние Кука

$$D_i = \frac{r_i^2}{p} \cdot \frac{h_i}{1 - h_i} \tag{35}$$

Интерпретация: $D_i > 1$ указывает на влиятельную точку данных.

7.5. Q-Q график

Сравнивает квантили стандартизированных остатков с теоретическими квантилями стандартного нормального распределения:

$$\Phi^{-1}\left(\frac{i - 0.5}{n}\right) \quad \text{vs.} \quad r_{(i)} \tag{36}$$

где $r_{(i)}$ – отсортированные стандартизированные остатки, а Φ^{-1} – квантильная функция стандартного нормального распределения.

8. Выбор модели

8.1. Информационный критерий Акаике (AIC)

$$\text{AIC} = n \ln(\text{RSS}/n) + 2k \tag{37}$$

где:

- $\text{RSS} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$: сумма квадратов остатков
- k: число оцениваемых коэффициентов регрессии (линейная, экспоненциальная, логарифмическая, степенная: $k = 2$; полином степени p: $k = p+1$; множественная регрессия с p переменными: $k = p+1$)

Принцип: выбирается модель с наименьшим AIC (баланс между качеством подгонки и сложностью модели). В сравнении моделей инструмент сортирует модели по R^2 и дополнительно показывает AIC.

9. Пример 1: функция потребления США 1960–2008

Инструмент содержит три экономических примера наборов данных (список выбора рядом с кнопкой *Загрузить пример*). Пример 1 подробно рассчитан здесь, примеры 2 и 3 — в следующих разделах. Все числа соответствуют результатам инструмента.

9.1. Набор данных

49 годовых значений 1960–2008: Потребление = реальные личные потребительские расходы, Доход = реальный располагаемый доход, ВВП = реальный валовой внутренний продукт (каждый в млрд долларов США в ценах 2005 г.), Ставка = ставка по 3-месячным казначейским векселям (%), Безработица = уровень безработицы (%), Инфляция (%). Источник: Federal Reserve Bank of St. Louis (FRED) и U.S. Bureau of Labor Statistics (набор данных «macrodata» пакета Python statsmodels), среднегодовые значения квартальных данных.

Год	Потребление	Доход	Ставка	Безработица
1960	1784,4	1964,0	2,71	5,58
1965	2241,8	2514,6	4,01	4,53
1970	2740,1	3108,5	6,11	5,00
1975	3214,1	3691,4	5,66	8,48
1980	3766,2	4295,0	11,68	7,18
1985	4540,4	5144,8	7,42	7,20
1990	5316,2	5896,3	7,38	5,60
1995	6079,0	6651,7	5,43	5,62
2000	7608,1	8161,6	5,80	3,95
2005	8819,0	9277,3	3,30	5,08
2008	9290,9	9911,1	1,15	5,80

Таблица 1: Выдержка из примера 1 (каждый пятый год; полные данные в инструменте)

9.2. Описательная статистика

	Потребление (Y)	Доход (X)
Среднее	4821,1	5307,9
Стандартное отклонение	2241,1	2340,0
Медиана	4315,6	4989,0
$Q_{0.25}$ / $Q_{0.75}$	3019,5 / 6291,2	3406,5 / 6871,0
IQR	3271,7	3464,5
Асимметрия	0,565	0,464
Эксцесс	$-0{,}739$	$-0{,}806$

Оба ряда слегка скошены вправо (среднее $>$ медианы): из-за роста более высокие значения последних лет дальше отстоят друг от друга.

9.3. Корреляционный анализ

При $n = 49$, $\sum x_i = 260085$, $\sum y_i = 236234$, $\sum x_i y_i = 1\,510\,639\,978{,}65$ и $\sum x_i^2 = 1\,648\,793\,107{,}72$:

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = 256\,743\,654{,}4, \qquad \sum (x_i - \bar{x})^2 = 268\,299\,082{,}7,$$
$$r_{XY} = 0{,}9992, \qquad \rho_{XY}\ (\text{Спирмен}) = 0{,}9998.$$

Интерпретация: почти идеальная линейная связь. Осторожно: для временных рядов с трендом высокая корреляция отчасти является следствием общего роста. Потребление слабо отрицательно коррелирует со ставкой ($r = -0{,}29$) и безработицей ($r = -0{,}14$).

9.4. Простая линейная регрессия: кейнсианская функция потребления

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{256\,743\,654{,}4}{268\,299\,082{,}7} = 0{,}9569,$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 4821{,}10 - 0{,}9569 \cdot 5307{,}86 = -258{,}15$$

Уравнение регрессии:

$$\boxed{\widehat{\text{Потребление}} = -258{,}15 + 0{,}9569 \cdot \text{Доход}} \tag{38}$$

- Наклон — это **предельная склонность к потреблению**: из каждого дополнительного доллара располагаемого дохода около 96 центов потребляется и около 4 центов сберегается.
- Средняя склонность к потреблению $\bar{y}/\bar{x} = 0{,}908$ ниже; поскольку $\hat{\beta}_0 < 0$, она слегка растёт с доходом.

Показатели качества:

$$\text{SST} = 246\,097\,042, \quad \text{SSE} = 411\,133, \quad R^2 = 1 - \frac{411\,133}{246\,097\,042} = 0{,}9983,$$
$$\text{RMSE} = \sqrt{411\,133/49} = 91{,}60, \quad \text{MAE} = 76{,}48 \quad (\text{млрд долларов США}).$$

Прогноз для дохода 10 000: $\hat{y} = -258{,}15 + 0{,}9569 \cdot 10\,000 = 9311{,}2$.

9.5. Анализ остатков

Стандартное отклонение остатков $s_e = 91{,}60$. Пример: 2005 год (наибольшее расстояние Кука):

$$\hat{y} = -258{,}15 + 0{,}9569 \cdot 9277{,}3 = 8619{,}58, \quad e = 8819{,}0 - 8619{,}58 = 199{,}42,$$
$$r = 199{,}42/91{,}60 = 2{,}18, \quad h = \tfrac{1}{49} + \tfrac{(9277{,}3 - 5307{,}86)^2}{268\,299\,082{,}7} = 0{,}079,$$
$$D = \tfrac{2{,}18^2}{2} \cdot \tfrac{0{,}079}{1-0{,}079} = 0{,}204.$$

Интерпретация:

- Лишь два года имеют $|r| > 2$ (1984: $-2{,}19$, 2005: $2{,}18$); все $D_i < 1$ — влиятельных выбросов нет.
- На графике остатков положительные и отрицательные остатки следуют друг за другом длительными фазами: остатки **автокоррелированы** (корреляция последовательных остатков 0,82, статистика Дарбина–Уотсона 0,35; оба значения инструмент не рассчитывает). Это типично для временных рядов — стандартное предположение о независимости ошибок нарушено, поэтому R^2 выглядит лучше, чем есть на самом деле.

9.6. Множественная регрессия

С доходом, ставкой и безработицей в качестве регрессоров:

$$\widehat{\text{Потребление}} = 56{,}84 + 0{,}9494 \cdot \text{Доход} - 18{,}21 \cdot \text{Ставка} - 30{,}17 \cdot \text{Безработица} \quad (39)$$

$R^2 = 0{,}99938$, $R^2_{\text{adj}} = 0{,}99933$, RMSE $= 56{,}00$ (вместо 91,60 в простой регрессии). VIF: 1,08; 1,22; 1,14 — мультиколлинеарности нет.

Интерпретация: при данном доходе ставка, выше на один процентный пункт, снижает потребление примерно на 18 млрд долларов США, а безработица, выше на один процентный пункт, — примерно на 30 млрд долларов США (сбережения из предосторожности). Предельная склонность к потреблению почти не меняется (0,95).

9.7. Сравнение моделей: рост во времени

При Y = Потребление и X = Год модель *Экспоненциальная*

$$\widehat{\text{Потребление}} = a \cdot e^{0{,}0340 \cdot \text{Год}}, \qquad R^2 = 0{,}9967, \quad (40)$$

показывает средний рост реального потребления $e^{0{,}034} - 1 \approx 3{,}5\,\%$ в год (a из-за использования календарных лет в качестве X чрезвычайно мало и не интерпретируется).

10. Пример 2: кривая Энгеля

40 домохозяйств (фиктивные, реалистичные значения): Доход = чистый доход домохозяйства; Питание, Жильё, Досуг = ежемесячные расходы (евро); Размер_домохозяйства (человек); Доля_питания (% дохода).

10.1. Степенная модель и эластичность по доходу

В степенной модели $y = a \cdot x^{b}$ параметр $b = \dfrac{\mathrm{d}\ln y}{\mathrm{d}\ln x}$ есть **эластичность по доходу**. Результаты при X = Доход (вкладка *Простая регрессия*):

Y	a	b	R^2	Классификация
Питание	14,84	0,412	0,718	товар первой необходимости ($0 < b < 1$)
Жильё	0,587	0,885	0,971	товар первой необходимости, близко к 1
Досуг	0,00903	1,296	0,960	предмет роскоши ($b > 1$)

Таблица 2: Эластичности по доходу (степенная модель)

Закон Энгеля: доля питания падает с ростом дохода ($r = -0{,}84$). Степенная модель для доли: $\widehat{\text{Доля_питания}} = 1478{,}9 \cdot \text{Доход}^{-0{,}588}$, $R^2 = 0{,}848$ — она подходит заметно лучше прямой ($R^2 = 0{,}706$), так как доля падает выпукло. Как и ожидалось, $-0{,}588 \approx 0{,}412 - 1$.

10.2. Множественная регрессия с размером домохозяйства

Для питания один лишь доход объясняет только 72 % дисперсии. С размером домохозяйства:

$$\widehat{\text{Питание}} = 91{,}32 + 0{,}0538 \cdot \text{Доход} + 46{,}47 \cdot \text{Размер_домохозяйства} \qquad (41)$$

$R^2 = 0{,}916$, $R^2_{\text{adj}} = 0{,}911$, VIF $= 1{,}05$. Каждый дополнительный человек при данном доходе увеличивает расходы на питание примерно на 46 евро в месяц; из каждого дополнительного евро дохода около 5 центов идёт на питание.

11. Пример 3: производственная функция Кобба–Дугласа

30 промышленных предприятий (фиктивные, реалистичные значения): Выпуск = добавленная стоимость, Капитал = основной капитал (каждый в тыс. евро), Занятые; кроме того, Производительность (выпуск на одного занятого), Капиталовооружённость (капитал на одного занятого) и натуральные логарифмы.

11.1. Линеаризация и множественная регрессия

$$Y = A\,K^{\alpha}L^{\beta} \quad \Longleftrightarrow \quad \ln Y = \ln A + \alpha \ln K + \beta \ln L \tag{42}$$

Множественная регрессия ln_Выпуск на ln_Капитал и ln_Занятые:

$$\widehat{\ln Y} = 0{,}933 + 0{,}3113 \cdot \ln K + 0{,}6920 \cdot \ln L, \qquad R^2 = 0{,}9982,\ R^2_{\text{adj}} = 0{,}9981. \tag{43}$$

- **Эластичности выпуска:** рост капитала на 1 % увеличивает выпуск на 0,31 %, рост числа занятых на 1 % — на 0,69 %.
- **Отдача от масштаба:** $\alpha + \beta = 1{,}003 \approx 1$ — постоянная отдача от масштаба.
- $A = e^{0{,}933} = 2{,}54$ (совокупная факторная производительность).
- **Мультиколлинеарность:** VIF $= 11{,}1$ (инструмент: «Высокая»), так как у крупных предприятий много капитала и много занятых ($r(\ln K, \ln L) = 0{,}95$). Здесь оценка остаётся устойчивой благодаря малым возмущениям; в реальных данных α и β из-за этого часто оцениваются неточно.

Смещение из-за пропущенной переменной: простая регрессия ln_Выпуск только на ln_Капитал даёт наклон 0,918 — намного выше $\alpha = 0{,}31$, поскольку ln_Капитал вбирает в себя влияние отсутствующих занятых.

11.2. Интенсивная форма

При постоянной отдаче от масштаба $Y/L = A\,(K/L)^{\alpha}$. Степенная модель при Y = Производительность, X = Капиталовооружённость:

$$\widehat{\text{Производительность}} = 2{,}568 \cdot \text{Капиталовооружённость}^{0{,}312}, \qquad R^2 = 0{,}862. \tag{44}$$

Показатель степени $0{,}312$ совпадает с α из множественной регрессии.

12. Итог: сравнение моделей

Вкладка *Сравнение моделей* для примера 1 при Y = Потребление, X = Доход (AIC как в инструменте, раздел «Выбор модели»):

Рекомендация: инструмент упорядочивает модели по R^2; при этом почти всегда побеждает полином наивысшей степени. AIC штрафует за дополнительные параметры; здесь полиномы начиная со степени 3 практически равноценны. Экономически осмысленны тем не менее линейная функция потребления (интерпретируемая предельная склонность к потреблению) или множественная регрессия со ставкой и безработицей, которая всего с четырьмя параметрами достигает значительной части улучшения. Полиномы высокой степени подстраиваются под выборку и непригодны для прогнозов за пределами диапазона данных.

Модель	R^2	RMSE	AIC
Линейная	0,99833	91,60	446,71
Полином степени 2	0,99952	49,15	387,69
Полином степени 3	0,99964	42,74	376,00
Полином степени 6	0,99968	40,39	376,47
Экспоненциальная	0,94932	504,51	613,91
Логарифмическая	0,93672	563,76	624,79
Степенная ($b = 1{,}035$)	0,99832	91,75	446,87
Множественная регрессия	0,99938	56,00	402,49

Таблица 3: Сравнение моделей регрессии для функции потребления

13. Литература

- Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression: Models, Methods and Applications*. Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. Wiley.
- Seber, G. A. F., & Wild, C. J. (2003). *Nonlinear Regression*. Wiley.
- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. Wiley.

14. Приложение: численные методы

14.1. Метод Гаусса

Для решения системы линейных уравнений $\mathbf{Ax} = \mathbf{b}$:
Алгоритм:

1. Прямой ход: приведение $\mathbf{A}$ к верхней треугольной форме
2. Обратный ход: решение относительно $\mathbf{x}$

14.2. Аппроксимация обратной функции ошибок

Для Q-Q графика требуется обратная функция ошибок erf^{-1}:

$$\operatorname{erf}^{-1}(x) \approx \operatorname{sgn}(x)\sqrt{\sqrt{\left(\frac{2}{\pi a}+\frac{\ln(1-x^2)}{2}\right)^2-\frac{\ln(1-x^2)}{a}}-\left(\frac{2}{\pi a}+\frac{\ln(1-x^2)}{2}\right)} \quad (45)$$

где $a = 0{,}147$ (аппроксимация Виницкого).

Теоретические квантили для Q-Q графика тогда:

$$q_i = \sqrt{2} \cdot \mathrm{erf}^{-1}\left(2 \cdot \frac{i - 0.5}{n} - 1\right), \quad i = 1, \ldots, n \tag{46}$$