2. Парная регрессия

2.1. Понятие регрессии

Парной регрессией называется уравнение связи двух переменных y и x вида

$$y = f(x), \tag{2.1}$$

где y – зависимая переменная (результативный признак); x – независимая, объясняющая переменная (признак-фактор).

Парная регрессия применяется, если имеется доминирующий фактор, который и используется в качестве объясняющей переменной.

Различают линейные и нелинейные относительно фактора x регрессии.

Наиболее часто применяются следующие модели регрессий:

– линейная – $\hat{y} = a + b \cdot x$;

– гиперболическая – $\hat{y} = a + b / x$;

– экспоненциальная – $\hat{y} = e^{a_0 + bx}$;

– модифицированная экспонента – $\hat{y} = K + a_0 \times b^x$;

– параболическая – $\hat{y} = a + b \cdot x + cx^2$;

– показательная – $\hat{y} = a \cdot b^x$;

– степенная – $\hat{y} = a \cdot x^b$;

– логарифмическая – $\hat{y} = a_0 + b \ln x$.

Интерпретация коэффициента b при факторной переменной x в линейной регрессии: коэффициент b показывает, на сколько единиц в среднем изменится величина переменной y при изменении фактора x на 1 единицу измерения.

2.2. Построение уравнения регрессии

Постановка задачи. По имеющимся данным n наблюдений за совместным изменением двух переменных x и y $\{(x_i,y_i),\ i=1,2,...,n\}$ (табл. 1.1) необходимо определить аналитическую зависимость $\hat{y}=f(x)$, наилучшим образом описывающую данные наблюдений.

Построение уравнения регрессии осуществляется в два этапа (предполагает решение двух задач):

– спецификация модели (определение вида аналитической зависимости $\hat{y}=f(x)$);

– оценка параметров (определение численных значений) выбранной модели.

2.2.1. Спецификация модели

Для выбора вида аналитической зависимости применяется три основных метода:

– графический (на основе анализа поля корреляций);

– аналитический, т. е. исходя из теории изучаемой взаимосвязи;

– экспериментальный, т. е. путем сравнения величины остаточной дисперсии $D_{\text{ост}}$ или средней ошибки аппроксимации $\overline{A}$, рассчитанных для различных моделей регрессии (метод перебора).

2.2.2. Оценка параметров линейной модели

Для оценки параметров регрессий, линейных по этим параметрам, используется *метод наименьших квадратов* (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических значений $\hat{y}_x$ при тех же значениях фактора x минимальна, т. е.

$$\sum\left(y-\hat{y}_x\right)^2 \to \min.$$

В случае *линейной регрессии* параметры a и b находятся из следующей системы нормальных уравнений метода МНК:

$$\begin{cases} na+b\sum x=\sum y, \\ a\sum x+b\sum x^2=\sum yx. \end{cases} \qquad (2.2)$$

Можно воспользоваться готовыми формулами, которые вытекают из этой системы:

$$a=\bar{y}-b\cdot\bar{x}, \qquad b=\frac{\operatorname{cov}(x,y)}{\sigma_x^2}=\frac{\overline{y\cdot x}-\bar{y}\cdot\bar{x}}{\overline{x^2}-\bar{x}^2}. \qquad (2.3)$$

2.2.3. Оценка параметров линейных моделей

Нелинейные уравнения регрессии предварительно приводятся к линейному виду с помощью преобразования переменных (обычно рассматриваются модели, для которых такое преобразование возможно), а затем к преобразованным переменным применяется обычный МНК. Для нелинейных уравнений регрессии, приводимых к линейным с помощью преобразования $(x, y) \to (x', y')$, система нормальных уравнений имеет вид (2.2) в преобразованных переменных x', y'. Далее для наиболее часто применяемых нелинейных моделей приводятся линеаризующие преобразования и формулы для расчета параметров.

Гиперболическая регрессия: $y_x = a_0 + a_1 / x$.

Линеаризующее преобразование: $x' = 1/x$; $y' = y$.

Уравнения (2.2) и формулы (2.3) принимают вид

$$\begin{cases} na+b\sum\frac{1}{x}=\sum y, \\ a\sum\frac{1}{x}+b\sum\frac{1}{x^2}=\sum y\frac{1}{x}. \end{cases} \qquad (2.4)$$

$$a_0=\frac{1}{n}\sum y-\frac{1}{n}a_1\sum\frac{1}{x}, \qquad a_1=\frac{n\sum\frac{y}{x}-\sum\frac{1}{x}\sum y}{n\sum\frac{1}{x^2}-\left(\sum\frac{1}{x}\right)^2}. \qquad (2.5)$$

Экспоненциальная регрессия: $y_x = e^{a_0+a_1x}$.

Линеаризующее преобразование и расчетные формулы: $x' = x$; $y' = \ln y$.

$$a_0 = \frac{1}{n}\sum \ln y - \frac{1}{n}a_1\sum x, \qquad a_1 = \frac{n\sum x\ln y - \sum x\sum \ln y}{n\sum x^2 - (\sum x)^2}. \tag{2.6}$$

Модифицированная экспонента: $y_x = K + a_0 \times a_1^{\ x}$, $(0 < a_1 < 1)$.

Линеаризующее преобразование и расчетные формулы: $x' = x$; $y' = ln\,|\,y - K\,|$.

$$\ln|a_0| = \frac{1}{n}\sum \ln|y-K| - \frac{1}{n}\ln a_1\sum x, \quad a_1 = \frac{n\sum(x\ln|y-K|) - \sum x\sum \ln|y-K|}{n\sum x^2 - (\sum x)^2}. \tag{2.7}$$

Величина предела роста K выбирается предварительно на основе анализа поля корреляций либо из качественных соображений. Параметр a_0 берется со знаком «+», если $y_x > K$ и со знаком «–» в противном случае.

Степенная функция: $y_x = a_0 \times x^{a_1}$, $(a_0 > 0)$.

Линеаризующее преобразование и расчетные формулы: $x' = ln\,x$; $y' = ln\,y$.

$$\ln a_0 = \frac{1}{n}\sum \ln y - \frac{1}{n}a_1\sum \ln x, \qquad a_1 = \frac{n\sum(\ln x\cdot \ln y) - \sum \ln x\sum \ln y}{n\sum(\ln x)^2 - (\sum \ln x)^2}. \tag{2.8}$$

Показательная функция: $y_x = a_0 \times a_1^x$.

Линеаризующее преобразование: $x' = x$; $y' = lny$.

$$\ln a_0 = \frac{1}{n}\sum \ln y - \frac{1}{n}\ln a_1\sum x, \qquad \ln a_1 = \frac{n\sum x\ln y - \sum x\sum \ln y}{n\sum x^2 - (\sum x)^2}. \tag{2.9}$$

Логарифмическая функция: $y_x = a_0 + a_1 \ln x$.

Линеаризующее преобразование: $x' = ln\,x$; $y' = y$.

$$a_0 = \frac{1}{n}\sum y - \frac{1}{n}a_1\sum \ln x, \qquad a_1 = \frac{n\sum \ln x\cdot y - \sum \ln x\sum y}{n\sum(\ln x)^2 - (\sum \ln x)^2}. \tag{2.10}$$

Парабола второго порядка: $y_x = a_0 + a_1 x + a_2 x^2$.

Линеаризующее преобразование: $x_1' = x$; $x_2' = x^2$; $y' = y$. Этой модели соответствуют две факторные переменные $x_1' = x$; $x_2' = x^2$.

Парабола второго порядка имеет 3 параметра a_0, a_1, a_2, которые определяются из системы трех уравнений

$$\begin{cases} n\cdot a_0 + a_1\sum x + a_2\sum x^2 = \sum y, \\ a_0\sum x + a_1\sum x^2 + a_2\sum x^3 = \sum xy, \\ a_0\sum x^2 + a_1\sum x^3 + a_2\sum x^4 = \sum x^2 y. \end{cases} \tag{2.11}$$

2.3. Оценка качества и точности построенной модели регрессии

Качество построенной модели регрессии оценивается с помощью индекса корреляции R (1.5) или *коэффициента детерминации* $R^2 = (R)^2$ вычисляемого как квадрат индекса корреляции. Для линейной регрессии $R^2 = r^2_{xy}$. Коэффициент детерминации R^2 принимает значения в диапазоне

$$0 \le R^2 \le 1.$$

Чем ближе величина R^2 к единице, тем лучше уравнение регрессии $\hat{y} = f(x)$ согласуется с данными наблюдений. При $R = 1$ зависимость $\hat{y} = f(x)$ становится функциональной, т. е. соотношение $\hat{y}_i = f(x_i)$ выполняется для всех наблюдений.

Коэффициент детерминации R^2 интерпретируется следующим образом. Величина R^2 показывает, какая доля общей дисперсии (вариации) результативного признака y объясняется уравнением регрессии. Например, значение $R^2 = 0{,}75$ означает, что уравнение регрессии объясняет 75% общей дисперсии (вариации) результативного признака y.

Точность построенной модели регрессии оценивается с помощью *средней квадратической ошибки* ($\varepsilon_{кв}$)

$$\varepsilon_{кв} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \tag{2.12}$$

либо *средней ошибки аппроксимации* ($\overline{A}$), которая представляет собой среднее относительное отклонение расчетных значений от наблюдаемых

$$\overline{A} = \frac{1}{n} \sum_{i=1}^{n} \frac{y - \hat{y}}{y} \cdot 100\,\%. \tag{2.13}$$

Чем выше ниже средняя ошибка аппроксимации $\varepsilon_{кв}$, тем лучше модель регрессии описывает исходные данные.

Построенное уравнение регрессии можно считать удовлетворительным, если значение $\overline{A}$ не превышает 10–12%.

2.4. Оценка значимости уравнения регрессии

Оценка статистической значимости уравнения регрессии в целом осуществляется с помощью F-критерия Фишера по следующему алгоритму:

1) выдвигается нулевая гипотеза $H_о$ о статистической незначимости уравнения регрессии (или коэффициента детерминации R^2);

2) вычисляется фактическое значение F–критерия $F_{факт}$ и определяется критическое (табличное) значение F–критерия $F_{табл}$;

3) проверяется условие $F_{факт} > F_{табл}$. Если условие выполняется, то нулевая гипотеза $H_о$ о статистической незначимости уравнения регрессии отвергается и уравнение считается статистически значимым. Если $F_{факт} \le F_{табл}$, то гипотеза $H_о$ не отклоняется и признается статистическая незначимость или ненадежность уравнения регрессии.

Величина $F_{факт}$ определяется по формуле

$$F_{ôàêò} = \frac{R^2}{1 - R^2} \cdot \frac{n - p - 1}{p}, \tag{2.14}$$

где n – число единиц совокупности; p – число параметров при факторных переменных. Для парной линейной регрессии $p = 1$.

$F_{табл}$ представляет собой табличное значение F-критерия Фишера при уровне значимости α и числе степеней свободы $k_1 = p$, $k_2 = n - p - 1$.

2.5. Оценка значимости коэффициентов уравнения регрессии

Для оценки статистической значимости коэффициентов линейной регрессии применяется t-критерий Стьюдента:

1) выдвигается нулевая гипотеза $H_о$ о статистической незначимости коэффициента уравнения регрессии;

2) вычисляется фактическое значение t–критерия $t_{факт}$ и определяется критическое (табличное) значение t–критерия $t_{табл}$;

3) проверяется условие $t_{факт} > t_{табл}$. Если условие выполняется, то нулевая гипотеза $H_о$ о статистической незначимости коэффициента уравнения регрессии отвергается и коэффициент уравнения считается статистически значимым. Если $t_{факт} \le t_{табл}$, то гипотеза $H_о$ не отклоняется и признается статистическая незначимость или ненадежность коэффициента уравнения регрессии.

Величины $t_{b,факт}, t_{a,факт}$ определяются по формулам

$$t_{b,факт} = \frac{b}{s_b}; \quad t_{a,факт} = \frac{a}{s_a}, \tag{2.15}$$

где s_a и s_b – стандартные ошибки коэффициентов регрессии (2.16), (2.17).

Величина $t_{крит} = t_{1-\alpha,n-2}$ представляет собой табличное значение t-критерия Стьюдента при уровне значимости α и числе степеней свободы $k = n - p - 1$ (определяется по таблицам).

2.6. Оценка точности коэффициентов уравнения регрессии

Получаемые оценки коэффициентов регрессии зависят от используемой выборки значений переменных x и y и являются случайными величинами. Представление о точности полученных оценок, о том, насколько далеко они могут отклониться от истинных значений коэффициентов, можно получить, используя, так называемые, «стандартные ошибки» коэффициентов регрессии.

Стандартные ошибки коэффициентов регрессии (s_a), (s_b) определяются соотношениями

$$s_b = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2/(n-2)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{s_{ост}^2}{\sum(x-\bar{x})^2}} = \frac{s_{ост}}{\sigma_x\sqrt{n}}, \tag{2.16}$$

$$s_a = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{(n-2)} \cdot \frac{\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{s_{ост}^2 \frac{\sum_{i=1}^{n}x_i^2}{n^2\sigma_x^2}} = s_{ост}\frac{\sqrt{\sum_{i=1}^{n}x_i^2}}{n\sigma_x}, \tag{2.17}$$

где $s^2_{ост}$ представляет собой несмещенную оценку остаточной дисперсии

$$s_{ост}^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{(n-2)}. \tag{2.18}$$

Проверка значимости оценок параметров ничего не говорит о том, насколько эти оценки могут отличаться от точных значений. Ответ на этот вопрос дает построение доверительных интервалов.

Под доверительным интервалом понимаются пределы, в которых лежит точное значение определяемого показателя с заданной вероятностью ($P = 1-\alpha$).

Доверительные интервалы для точных значений параметров $\widetilde{a}$ и $\widetilde{b}$ уравнения линейной регрессии определяются соотношениями:

$$a - t_{1-\alpha,n-2} \cdot s_a < \widetilde{a} < a + t_{1-\alpha,n-2} \cdot s_a; \quad b - t_{1-\alpha,n-2} \cdot s_b. < \widetilde{b} < b + t_{1-\alpha,n-2} \cdot s_b \qquad (2.19)$$

Величина $t_{1-\alpha,n-2}$ представляет собой табличное значение t-критерия Стьюдента при уровне значимости α и числе степеней свободы n–2 (п. 2.5).

Если в границы доверительного интервала попадает ноль, т. е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается равным нулю, так как он не может одновременно принимать и положительное, и отрицательное значения.

2.7. Точечный и интервальный прогноз по уравнению линейной регрессии

Точечный прогноз заключается в получении прогнозного значения y_p, которое определяется путем подстановки в уравнение регрессии $\hat{y}_x = a + b \cdot x$ соответствующего (прогнозного) значения x_p

$$y_p = a + b \cdot x_p.$$

Интервальный прогноз заключается в построении доверительного интервала прогноза, т. е. нижней и верхней границ y_{pmin} , y_{pmax} интервала, содержащего точную величину для прогнозного значения y_p ($y_{p\min} < y_p < y_{p\min}$) с заданной вероятностью.

При построении доверительного интервала прогноза используется стандартная ошибка индивидуального значения прогноза s_{y_p}, связанная с дисперсией ошибки прогноза $s^2_{y_p}$ соотношением $s_{y_p} = \sqrt{s^2_{y_p}}$.

$$s_{y_p} = s_{ост} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \overline{x})^2}{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2}} . \qquad (2.20)$$

Доверительный интервал для индивидуального значения прогноза y_p определяется соотношением:

$$y_p - t_{1-\alpha;n-2} \cdot s_{y_p} \le \hat{y}_p \le y_p + t_{1-\alpha;n-2} \cdot s_{y_p}, \qquad (2.21)$$

где величина $t_{1-\alpha,n-2}$ представляет собой табличное значение t-критерия Стьюдента на уровне значимости α при числе степеней свободы n–2.

2.8. Коэффициент эластичности

В экономических исследованиях широкое применение находит такой показатель, как *коэффициент эластичности* (Э), вычисляемый по формуле

$$Э = f'(x)\frac{x}{y}. \tag{2.19}$$

Коэффициент эластичности показывает, на сколько процентов изменится результат *y* при изменении фактора *x* на 1% от своего номинального значения. Для линейной регрессии коэффициент эластичности равен

$$Э = b\frac{x}{y} \tag{2.20}$$

и зависит от *x*, поэтому рассчитывают средний коэффициент эластичности

$$\overline{Э} = f'(\overline{x})\frac{\overline{x}}{\overline{y}} = b\frac{\overline{x}}{\overline{y}}. \tag{2.21}$$

Средний коэффициент эластичности $(\overline{Э})$ показывает, на сколько процентов в среднем по совокупности изменится результат *y* от своей величины при изменении фактора *x* на 1% от своего номинального значения.

Контрольные вопросы

1. Что понимается под парной регрессией?
2. Какие задачи решаются при построении уравнения регрессии?
3. Какие методы применяются для выбора вида модели регрессии?
4. Какие функции чаще всего используются для построения уравнения парной регрессии?
5. Какой вид имеет система нормальных уравнений метода наименьших квадратов в случае линейной регрессии?
6. Как вычисляется и что показывает индекс детерминации?
7. Как проверяется значимость уравнения регрессии?
8. Как проверяется значимость коэффициентов уравнения регрессии?
9. Понятие доверительного интервала для коэффициентов регрессии.
10. Понятие точечного и интервального прогноза по уравнению линейной регрессии.
11. Как вычисляются и что показывают коэффициент эластичности Э, средний коэффициент эластичности $\overline{Y}$?

Задачи

1. Из предложенных уравнений регрессии выбрать лучшее, т. е. то, которое дает лучшее приближение к данным наблюдения

$\hat{y}_x = 21,5 + 4,35 \cdot x$, $R^2 = 0,95$,

$\hat{y}_x = 20 + 1,115 \cdot \ln(x)$, $R^2 = 0,79$.

(Первое)

2. По величине коэффициента детерминации $R^2 = 0,56$ определить долю вариации результативного признака, объясненного уравнением регрессии. (56%)

3. Найти критические значения *F*–критерия и *t*–критерия по количеству наблюдений и уровню значимости: $n = 50$, $\alpha = 0,01$, $m = 1$; $n = 20$, $\alpha = 0,05$, $m = 1$, где m – количество факторных переменных. (7,19; 4,41).

4. По величине коэффициента детерминации R^2 = 0,4 проверить значимость (α=0,05) уравнения линейной парной регрессии. Число наблюдений n = 50. (Значимо).

5. По заданному уравнению регрессии

$$\hat{y}_x = 21{,}5 + 4{,}35 \cdot x,$$

найти средний коэффициент эластичности, если $\bar{x} = 52, \bar{y} = 250$. (0,90).

6. По заданному коэффициенту эластичности Э = 1,5 определить, на сколько изменится y при изменении x на 2 единицы, если до изменения признаки y и x принимали значения x = 40, y = 10. (0,75).

Лабораторная работа №2. Парный регрессионный анализ: построение модели в виде парной регрессии и проверка ее качества

Задание. На основании данных таблицы П1.2 для соответствующего варианта (табл. 2.1):

1. Построить предложенные в таблице 2.1 уравнения регрессии, включая линейную регрессию, используя формулы (2.3) – (2.11).
2. Вычислить показатели качества и точности для каждого уравнения.
3. Проверить значимость уравнений регрессии при уровнях значимости 0,05 и 0,01.
4. Определить лучшее уравнение регрессии на основе средней ошибки аппроксимации.
5. Проверить значимость коэффициентов линейной регрессии и построить доверительные интервалы для точных значений параметров $\tilde{a}$ и $\tilde{b}$ уравнения линейной регрессии с уровнем значимости 0,05.
6. Построить точечный и интервальный прогноз для значения $x = x_{max}$ по уравнению линейной регрессии с уровнем значимости 0,05.
7. Определить средний коэффициент эластичности по уравнению линейной регрессии.
8. Графически представить результаты моделирования.

Таблица 2.1

Варианты кривых выравнивания к лабораторной работе №2

Вариант	Графы из табл. П1.2 (x, y)	Виды кривых выравнивания					
		Линейная	Степенная	Экспоненциальная	Показательная	Логарифмическая	Гиперболическая
1	1,14	*	*				
2	2,14	*		*			
3	4,14	*			*		
4	6,14	*				*	
5	9,14	*					*
6	11,14	*	*				
7	12,14	*		*			

Окончание таблицы 2.1

Вариант	Графы из табл. П1.2	Виды кривых выравнивания					
		Линейная	Степенная	Экспоненциальная	Показательная	Логарифмическая	Гиперболическая
8	2,15	*			*		
9	3,15	*				*	
10	7,15	*					*
11	8,15	*	*				
12	12,15	*		*			
13	1,17	*			*		
14	2,17	*				*	
15	4,17	*					*
16	6,17	*	*				
17	9,17	*		*			
18	11,17	*			*		
19	12,17	*				*	
20	1,19	*					*
21	2,19	*	*				
22	4,19	*		*			
23	6,19	*			*		
24	9,19	*				*	
25	11,19	*					*

Пример выполнения лабораторной работы №2

Исходные данные:

– наблюдаемые значения переменных x и y заданы в таблице 2.2;
– построить две модели: линейную и степенную.

Таблица 2.2

Исходные данные для примера выполнения лабораторной работы №2

	Области	x	y		Области	x	y
1	Белгородская	113	39	12	Рязанская область	120	34
2	Брянская	124	37	13	Смоленская	125	39
3	Владимирская	124	36	14	Тамбовская	118	37
4	Воронежская	122	36	15	Тверская	122	35
5	Ивановская	128	26	16	Тульская	133	54
6	Калужская	140	43	17	Ярославская	136	36
7	Костромская	117	31	18	Архангельская	136	35
8	Курская	113	40	19	Вологодская	138	34
9	Липецкая	122	48	20	Калининградская	124	48
10	Московская	139	64	21	Ленинградская	123	30
11	Орловская	126	39	22	Мурманская	149	59

1) Определим коэффициенты a и b *линейной регрессии* (2.3), используя результаты промежуточных расчетов, приведенные в таблице 2.3.

Таблица 2.3

Промежуточные результаты расчетов для линейной регрессии

Номер наблюдения	x	y	x^2	y^2	xy	$\hat{y}$	$\lvert(\hat{y}-y)/y\rvert$	$(\hat{y}-y)^2$	$(y-\bar{y})^2$
1	113	39	12769	1521	4407	33,381	0,150	31,57	1,00
2	124	37	15376	1369	4588	38,616	0,007	2,61	9,00
3	124	36	15376	1296	4464	38,616	0,034	6,84	16,00
4	122	36	14884	1296	4392	37,664	0,014	2,77	16,00
5	128	26	16384	676	3328	40,519	0,490	210,81	196,00
6	140	43	19600	1849	6020	46,230	0,008	10,43	9,00
7	117	31	13689	961	3627	35,284	0,117	18,36	81,00
8	113	40	12769	1600	4520	33,381	0,171	43,81	0,00
9	122	48	14884	2304	5856	37,664	0,240	106,84	64,00
10	139	64	19321	4096	8896	45,754	0,329	332,92	576,00
11	126	39	15876	1521	4914	39,567	0,026	0,32	1,00
12	120	34	14400	1156	4080	36,712	0,051	7,36	36,00
13	125	39	15625	1521	4875	39,092	0,035	0,01	1,00
14	118	37	13924	1369	4366	35,760	0,054	1,54	9,00
15	122	35	14884	1225	4270	37,664	0,043	7,10	25,00
16	133	54	17689	2916	7182	42,899	0,247	123,24	196,00
17	136	36	18496	1296	4896	44,326	0,161	69,33	16,00
18	136	35	18496	1225	4760	44,326	0,194	86,98	25,00
19	138	34	19044	1156	4692	45,278	0,252	127,19	36,00
20	124	48	15376	2304	5952	38,616	0,224	88,07	64,00
21	123	30	15129	900	3690	38,140	0,229	66,26	100,00
22	149	59	22201	3481	8791	50,513	0,206	72,04	361,00
Сумма	2792	880	356192	37038	112566	880	3,282	1416,371	1838
Среднее значение	126,91	40	16190,55	1683,545	5116,636	40	0,149	64,381	83,545

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{5116{,}636 - 126{,}91 \cdot 40}{16190{,}55 - 126{,}91^2} = 0{,}476,$$

$$a = \bar{y} - b \cdot \bar{x} = 40 - 126{,}91 \cdot 0.476 = -20{,}39.$$

Уравнение линейной регрессии: $y = -20{,}39 + 0{,}476 \cdot x$.

2) Для построения *степенной модели* $y_x = a_0 \times x^{a_1}$ введем новые переменные $x' = \ln x$; $y' = \ln y$, вычислим значения новых переменных и выполним промежуточные расчеты (табл. 2.4).

Таблица 2.4

Промежуточные результаты расчетов для степенной регрессии

Номер наблюдения	lnx	lny	ln^2x	ln^2y	$lnx \cdot lny$	$\hat{y}$	$\lvert(\hat{y}-y)/y\rvert$	$(\hat{y}-y)^2$	$(y-\bar{y})^2$
1	4,727	3,664	22,348	13,422	17,319	33,850	0,144	26,523	1,000
2	4,820	3,611	23,235	13,039	17,406	38,065	0,044	1,134	9,000
3	4,820	3,584	23,235	12,842	17,274	38,065	0,073	4,265	16,000
4	4,804	3,584	23,079	12,842	17,215	37,291	0,046	1,667	16,000
5	4,852	3,258	23,542	10,615	15,808	39,623	0,558	185,585	196,000
6	4,942	3,761	24,420	14,147	18,587	44,373	0,075	1,884	9,000
7	4,762	3,434	22,678	11,792	16,353	35,371	0,138	19,103	81,000
8	4,727	3,689	22,348	13,608	17,439	33,850	0,165	37,823	0,000
9	4,804	3,871	23,079	14,986	18,597	37,291	0,215	114,681	64,000
10	4,934	4,159	24,349	17,296	20,522	43,973	0,285	401,095	576,000
11	4,836	3,664	23,390	13,422	17,718	38,842	0,015	0,025	1,000
12	4,787	3,526	22,920	12,435	16,882	36,520	0,080	6,353	36,000
13	4,828	3,664	23,313	13,422	17,689	38,453	0,002	0,299	1,000
14	4,771	3,611	22,759	13,039	17,227	35,753	0,034	1,555	9,000
15	4,804	3,555	23,079	12,640	17,080	37,291	0,076	5,249	25,000
16	4,890	3,989	23,916	15,912	19,508	41,588	0,206	154,048	196,000
17	4,913	3,584	24,134	12,842	17,605	42,777	0,231	45,928	16,000
18	4,913	3,555	24,134	12,640	17,466	42,777	0,266	60,483	25,000
19	4,927	3,526	24,278	12,435	17,375	43,573	0,332	91,649	36,000
20	4,820	3,871	23,235	14,986	18,660	38,065	0,196	98,702	64,000
21	4,812	3,401	23,157	11,568	16,367	37,678	0,271	58,947	100,000
22	5,004	4,078	25,039	16,626	20,404	48,007	0,144	120,855	361,000
Сумма	106,50	80,638	515,667	296,556	390,500	863,077	3,596	1437,853	1838,000
Среднее значение	4,841	3,665	23,439	13,480	17,750	39,231	0,163	65,357	83,545

Используя формулы (2.8), получим

$$\ln a_0 = \frac{1}{n}\sum \ln y - \frac{1}{n} a_1 \sum \ln x, \qquad a_1 = \frac{n\sum(\ln x \cdot \ln y) - \sum \ln x \sum \ln y}{n\sum(\ln x)^2 - (\sum \ln x)^2}.$$

Вычислим значения параметров

$$a_1 = \frac{\overline{\ln x \cdot \ln y} + \overline{\ln x} \cdot \overline{\ln y}}{\overline{\ln^2 x} - \overline{\ln x}^2} = \frac{39{,}231 - 4{,}841 \cdot 3{,}665}{23{,}439 - 4{,}841^2} = 1{,}263,$$

$$\ln a_0 = \overline{\ln y} - a_1 \overline{\ln x} = 3{,}665 - 1{,}263 \cdot 4{,}841 = 2{,}451,$$

$$a_0 = \exp(2{,}451) = 0{,}0862.$$

Уравнение степенной регрессии имеет вид: $y = 0{,}0862 \cdot x^{1{,}263}$.

3) Вычисление показателей качества: индекс корреляции R (1.5), коэффициент детерминации R^2, средняя квадратическая ошибка $\varepsilon_{\text{кв}}$ (2.12), средняя ошибки аппроксимации $\overline{A}$ (2.13).

Для *линейной* модели (табл. 2.3)

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}} = \sqrt{1 - \frac{1416{,}371}{1838}} = 0{,}479\,,$$

$$R^2 = R \cdot R = 0{,}479 \cdot 0{,}479 = 0{,}229,$$

$$\varepsilon_{\text{кв}} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} = \sqrt{\frac{1416{,}371}{22}} = \sqrt{64{,}381} = 8{,}024\,,$$

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\frac{y - \hat{y}}{y} \cdot 100\,\% = \frac{1}{22}3{,}282 \cdot 100\% = 14{,}9\%\,.$$

Для *степенной* модели (табл. 2.4)

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}} = \sqrt{1 - \frac{1437{,}853}{1838}} = 0{,}467\,,$$

$$R^2 = R \cdot R = 0{,}467 \cdot 0{,}467 = 0{,}218,$$

$$\varepsilon_{\text{кв}} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} = \sqrt{\frac{1437{,}853}{22}} = \sqrt{65{,}357} = 8{,}084\,,$$

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\frac{y - \hat{y}}{y} \cdot 100\,\% = \frac{1}{22}3{,}596 \cdot 100\% = 16{,}3\%\,.$$

4) Проверка значимости уравнений регрессии (п. 2.4).

Для *линейной* модели (табл. 2.3):

$$F_{\text{факт}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} = \frac{0{,}229}{1 - 0{,}229} \cdot \frac{22 - 1 - 1}{1} = 5{,}95\,.$$

$F_{\text{крит},0,05}$ = FРАСПОБР(0,05;1;20)=4,35.

$F_{\text{крит},0,01}$ = FРАСПОБР(0,01;1;20)=8,10.

При $\alpha = 0{,}05$ линейное уравнение значимо, при $\alpha = 0{,}01$ – не значимо.

Для *степенной* модели (табл. 2.4):

$$F_{\text{факт}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} = \frac{0{,}218}{1 - 0{,}218} \cdot \frac{22 - 1 - 1}{1} = 5{,}57\,.$$

$F_{\text{крит},0,05}$ и $F_{\text{крит},0,01}$ – те же самые.

При $\alpha = 0{,}05$ степенное уравнение значимо, при $\alpha = 0{,}01$ – не значимо.

5) Определение лучшего уравнения регрессии (по средней ошибке аппроксимации).

Так как $\overline{A}_{лин} = 14{,}9\% > \overline{A}_{степ} = 16{,}3\%$, то линейная модель дает меньшую погрешность.

6) Проверка значимости коэффициентов линейной регрессии (п. 2.5).

Определим оценку остаточной дисперсии $s^2_{ост}$, используя данные таблицы 2.3

$$s^2_{ост} = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{(n-2)} = \frac{1416{,}37}{22-2} = 70{,}8\,.$$

Определим стандартные ошибки коэффициентов регрессии (s_a), (s_b), используя значение σ_x, полученное в лабораторной работе №1:

$$s_b = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2/(n-2)}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = \frac{\sqrt{s^2_{ост}}}{\sigma_x\sqrt{n}} = \frac{\sqrt{70{,}8}}{22\cdot 9{,}199} = 0{,}0416\,,$$

$$s_a = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{(n-2)} \cdot \frac{\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}(x_i - \overline{x})^2}} = \sqrt{s^2_{ост}} \cdot \frac{\sqrt{\sum_{i=1}^{n}x_i^2}}{n\sigma_x} = \sqrt{70{,}8} \cdot \frac{\sqrt{356192}}{22\cdot 9{,}199} = 24{,}81\,.$$

Вычислим (2.15)

$$t_a = \frac{a}{s_a} = \frac{-20{,}39}{24{,}81} = -0{,}82, \qquad t_b = \frac{b}{s_b} = \frac{0{,}476}{0{,}0416} = 11{,}44\,.$$

Значение $t_{1-\alpha,n-2}$ определим по таблице П4.2 из приложения.

При $\alpha = 0{,}05$ и степени свободы $k = n–2 = 20–2 = 20$ получаем $t_{1-\alpha,n-2} = 2{,}09$.

Так как $|t_a| = 0{,}82 < t_{1-\alpha,n-2} = 2{,}09$, то делаем вывод о *незначимости* коэффициента a.

Так как $|t_b| = 11{,}44 > t_{1-\alpha,n-2} = 2{,}09$, то делаем вывод о *значимости* коэффициента b.

7) Определение доверительных интервалов для точных значений параметров $\tilde{a}$ и $\tilde{b}$ уравнения линейной регрессии (2.16) – (2.19).

Для точного значения параметра $\tilde{a}$ доверительный интервал определять не нужно, так как значение коэффициента a не значимо.

Доверительный интервал для точного значения параметра $\tilde{b}$

(0,476 – 2,09·0,0416; 0,476 + 2,09·0,0416) = (0,389; 0,563).

8) Построение точечного и интервального прогноза для значения $x = x_{\max}$ по уравнению линейной регрессии.

$$x_{\max} = 149.$$

Точечный прогноз y_p

$$y_p = -20{,}39 + 0{,}476\cdot x_{\max} = -20{,}39 + 0{,}476\cdot 149 = 50{,}53.$$

Вычислим

$$s_{y_p} = s_{ост} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{\max} - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}} = \sqrt{70,8} \cdot \sqrt{1 + \frac{1}{22} + \frac{(149 - 126,9)^2}{1861,7}} = 9,62 .$$

Интервальный прогноз (2.21)

$$((y_p - t_{1-\alpha;n-2} \cdot s_{y_p} \;;\; y_p + t_{1-\alpha;n-2} \cdot s_{y_p}) =$$

= (50,53 – 2,09·9,62; 50,53 +2,09·9,62) = (30,4; 70,6)

8) Определение среднего коэффициента эластичности по уравнению линейной регрессии (2.17).

$$\overline{Y} = b \frac{\overline{x}}{\overline{y}} = 0,476 \frac{126,91}{40} = 1,51.$$

Результаты:

1) Уравнение линейной регрессии $y = -20,39 + 0,476 \cdot x$.

2) Уравнение степенной регрессии $y = 0,0862 \cdot x^{1,263}$.

3) Показатели качества и точности:

Для *линейной* модели (табл. 2.3)

$R = 0,479$,
$R^2 = 0,229$,
$\varepsilon_{кв} = 8,024$,
$\overline{A} = 14,9\%$.

Для степенной модели (табл. 2.4)

$R = 0,467$,
$R^2 = 0,218$,
$\varepsilon_{кв} = 8,084$,
$\overline{A} = 16,3\%$.

4) Линейное уравнение значимо при α = 0,05 и не значимо при α = 0,01.
Степенное уравнение значимо при α = 0,05 и не значимо при α = 0,01.

5) Линейная модель дает меньшую погрешность.

6) Коэффициент *a незначим*,
коэффициент *b значим*.

7) Доверительный интервал для точного значения параметра $\tilde{b}$
(0,389; 0,563).

8) Точечный прогноз y_p = 50,53.
Интервальный прогноз (30,4; 70,6).

9) Средний коэффициент эластичности $\overline{Y} = 1,51$.

10) Графическое представление результатов моделирования (рис. 2.1).

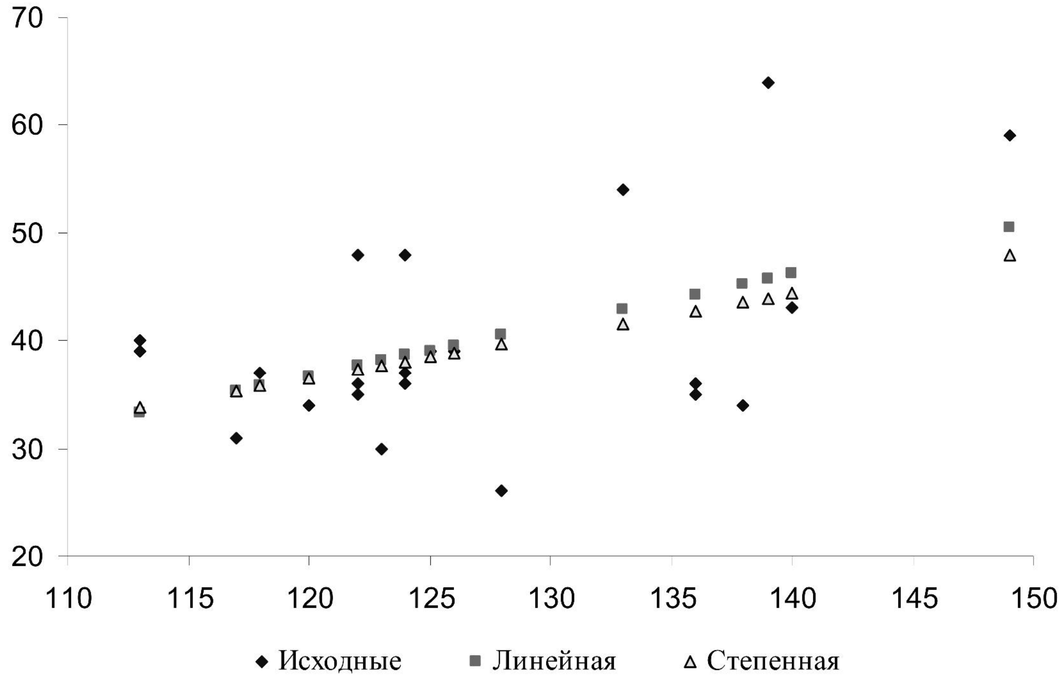

Рис. 2.1. Графическое представление результатов моделирования.

Таблица П1.2

Исходные данные к лабораторным работам №2, 3

Факторные переменные												
1	2	3	4	5	6	7	8	9	10	11	12	13
0,16	0,11	2,40	0,16	14,99	0,80	0,57	12,01	0,81	74,96	3,35	2,39	50,46
0,80	0,19	5,44	0,24	14,72	4,01	0,96	27,20	1,20	73,61	16,83	4,02	114,26
0,94	0,20	5,87	0,32	4,55	4,72	1,00	29,33	1,58	22,77	19,80	4,19	123,17
0,26	0,15	9,65	0,48	11,57	1,29	0,75	48,27	2,39	57,86	5,44	3,13	202,72
0,27	0,05	8,11	0,13	1,64	1,36	0,23	40,53	0,66	8,18	5,73	0,99	170,22
0,47	0,05	8,23	0,10	16,75	2,35	0,25	41,17	0,50	83,73	9,87	1,04	172,90
0,34	0,16	3,89	0,22	17,56	1,71	0,80	19,47	1,12	87,81	7,17	3,35	81,79
0,31	0,06	7,97	0,13	18,92	1,55	0,28	39,87	0,66	94,60	6,53	1,18	167,45
0,65	0,12	0,97	0,11	6,11	3,27	0,61	4,83	0,53	30,57	13,74	2,55	20,29
0,06	0,08	9,05	0,30	14,92	0,31	0,40	45,24	1,48	74,62	1,28	1,66	190,01
0,14	0,05	9,20	0,18	19,72	0,70	0,25	46,02	0,90	98,58	2,95	1,05	193,29
0,10	0,09	5,00	0,14	19,90	0,52	0,46	25,01	0,68	99,52	2,20	1,94	105,02
0,81	0,15	5,69	0,24	9,36	4,04	0,75	28,47	1,18	46,81	16,95	3,14	119,58
0,19	0,11	7,32	0,36	3,61	0,93	0,53	36,59	1,81	18,07	3,89	2,22	153,66
0,39	0,05	2,99	0,17	1,02	1,97	0,26	14,93	0,85	5,08	8,26	1,10	62,71
0,91	0,11	3,05	0,19	3,11	4,55	0,55	15,27	0,94	15,53	19,12	2,32	64,15
0,64	0,02	4,17	0,48	1,89	3,20	0,11	20,87	2,39	9,45	13,44	0,46	87,66
0,87	0,01	9,54	0,03	9,53	4,33	0,06	47,72	0,16	47,64	18,17	0,24	200,41
0,33	0,14	6,88	0,15	0,01	1,64	0,69	34,41	0,76	0,03	6,90	2,92	144,51
0,92	0,16	8,51	0,04	11,08	4,62	0,78	42,57	0,20	55,39	19,39	3,26	178,80
0,49	0,11	0,94	0,09	2,98	2,44	0,53	4,69	0,45	14,92	10,24	2,24	19,70
0,17	0,01	7,51	0,10	1,88	0,83	0,03	37,56	0,52	9,39	3,48	0,14	157,75
0,47	0,07	0,81	0,17	7,65	2,33	0,35	4,06	0,84	38,25	9,80	1,45	17,06
0,79	0,15	5,16	0,44	0,02	3,97	0,75	25,80	2,18	0,08	16,67	3,17	108,35
0,22	0,08	6,21	0,33	4,25	1,08	0,38	31,06	1,66	21,25	4,53	1,60	130,45
0,39	0,04	9,38	0,47	0,60	1,95	0,21	46,88	2,34	3,01	8,20	0,86	196,90
0,57	0,04	4,28	0,10	2,30	2,84	0,22	21,38	0,52	11,51	11,94	0,93	89,80
0,24	0,05	3,42	0,30	10,11	1,20	0,24	17,10	1,52	50,53	5,03	1,00	71,83
0,08	0,20	3,90	0,06	0,10	0,39	1,00	19,52	0,28	0,48	1,64	4,19	81,98
0,53	0,08	4,38	0,11	17,98	2,66	0,40	21,90	0,56	89,90	11,18	1,69	91,97
0,24	0,11	5,30	0,28	1,34	1,19	0,55	26,51	1,41	6,70	5,01	2,29	111,35
0,12	0,13	1,63	0,39	6,40	0,58	0,65	8,15	1,95	32,01	2,45	2,74	34,25
0,76	0,09	5,71	0,47	1,86	3,78	0,44	28,53	2,37	9,32	15,87	1,85	119,84
0,61	0,01	7,65	0,45	3,49	3,07	0,03	38,25	2,24	17,47	12,90	0,12	160,64
0,85	0,13	0,82	0,41	15,02	4,23	0,63	4,10	2,05	75,10	17,78	2,63	17,24
0,39	0,20	4,50	0,38	10,15	1,95	0,99	22,50	1,89	50,76	8,17	4,18	94,51
0,23	0,17	1,17	0,09	14,31	1,17	0,85	5,87	0,47	71,55	4,90	3,58	24,66
0,77	0,10	5,71	0,28	6,39	3,85	0,51	28,53	1,42	31,96	16,15	2,15	119,82
0,29	0,15	8,93	0,48	11,19	1,46	0,76	44,64	2,38	55,95	6,13	3,20	187,48
0,99	0,15	1,63	0,12	0,30	4,97	0,76	8,17	0,59	1,51	20,89	3,19	34,32
0,83	0,17	8,58	0,08	17,06	4,17	0,86	42,91	0,42	85,30	17,50	3,61	180,22
0,44	0,13	4,19	0,46	1,50	2,19	0,65	20,96	2,28	7,50	9,20	2,73	88,02
0,03	0,16	6,48	0,34	12,22	0,14	0,81	32,40	1,69	61,08	0,60	3,38	136,08
0,07	0,11	2,37	0,34	5,36	0,33	0,57	11,85	1,71	26,79	1,41	2,40	49,77
0,75	0,16	2,80	0,10	3,24	3,74	0,80	14,00	0,50	16,22	15,69	3,36	58,78

Окончание таблицы П1.2

Переменная y					
14	15	16	17	18	19
21,1	15,0	-0,6	20,7	23,0	12,7
20,6	10,3	43,0	20,5	24,4	45,7
20,7	4,5	49,2	20,6	25,7	51,0
22,1	-0,5	51,7	21,0	28,5	101,7
20,2	-4,5	27,6	19,9	20,3	49,5
19,8	2,9	41,7	19,6	19,4	62,4
21,3	14,2	26,2	20,8	24,8	34,8
20,2	4,6	23,7	20,0	20,4	41,4
20,0	13,6	24,2	19,9	20,6	5,9
21,4	0,9	36,8	20,6	23,4	80,1
20,7	2,4	42,4	20,2	21,1	83,9
20,9	11,8	32,0	20,6	22,3	51,6
20,3	6,3	36,4	20,2	22,7	38,9
21,6	-0,7	17,4	20,7	24,8	42,5
20,1	5,6	33,0	19,8	20,2	30,8
19,6	7,7	45,1	19,6	20,2	29,7
20,3	3,0	31,6	19,8	18,8	38,6
18,5	-4,1	70,8	18,4	16,8	94,3
20,9	-1,0	26,3	20,6	23,0	44,8
19,5	1,6	71,4	19,5	19,4	88,9
20,1	11,8	23,8	20,0	20,7	6,2
20,0	-3,9	34,6	19,8	19,5	64,1
20,1	13,6	4,7	19,9	20,4	3,6
20,9	2,7	25,3	20,7	25,8	30,4
21,2	1,2	31,1	20,4	22,9	57,3
21,0	-7,6	46,4	20,1	21,0	83,4
19,5	3,5	19,9	19,4	18,9	23,3
20,8	9,2	27,5	20,1	21,2	36,4
21,6	6,2	1,0	21,5	24,0	18,0
19,9	11,8	19,6	19,8	20,1	24,5
21,2	2,2	30,2	20,6	23,9	50,3
22,0	12,5	1,4	20,9	26,7	14,7
20,6	1,3	40,0	20,3	22,6	49,6
20,2	-3,4	54,4	19,6	17,9	81,0
20,5	18,4	29,6	20,4	23,8	10,7
21,9	10,1	22,8	21,3	29,1	36,9
21,2	18,2	-2,8	21,0	23,4	4,0
20,1	3,8	50,4	19,9	21,5	57,1
22,1	0,8	49,5	21,0	28,5	93,8
19,6	9,9	37,4	19,6	20,2	12,4
20,0	4,8	44,1	19,9	20,9	47,8
21,5	5,0	28,8	20,8	26,2	40,9
22,2	6,4	18,1	21,3	27,9	48,8
21,8	10,2	20,0	20,8	25,5	27,2
20,1	9,2	14,5	20,0	21,2	11,5