

Однофакторные регрессионные модели: понятие, виды и параметры

Задачей моделирования и прогнозирования с использованием регрессионной модели является оценка результата y (или результирующей переменной) на основе объясняющей переменной x . **Однофакторная (простая, парная) регрессия** – частный случай многофакторной регрессии, в которой только одна объясняющая переменная как в классической функциональной зависимости:

$$y = f(x).$$

Однофакторная Регрессия (*simple regression model*) – это регрессия между результирующей y и одной объясняющей переменной x . В общем виде её можно задать

$$y = \hat{f}(x) = f(x) + e,$$

где e – дополнительный остаточный член.

Графически построение однофакторной регрессии можно показать на рис. 4.1, разность между фактическим и теоретическим значением обозначена стрелочками, это ошибка регрессии

$$e = \hat{f}(x) - f(x) = \hat{y} - y.$$

Примеры однофакторной регрессии: зависимость производительности труда персонала от фондовооруженности.

Однофакторные регрессии могут быть динамические (будут рассматриваться в пятой главе) и пространственные.

Однофакторные регрессии можно разделить на линейные и нелинейные. При этом многие нелинейные функции можно привести к линейному виду (табл. 3.1).

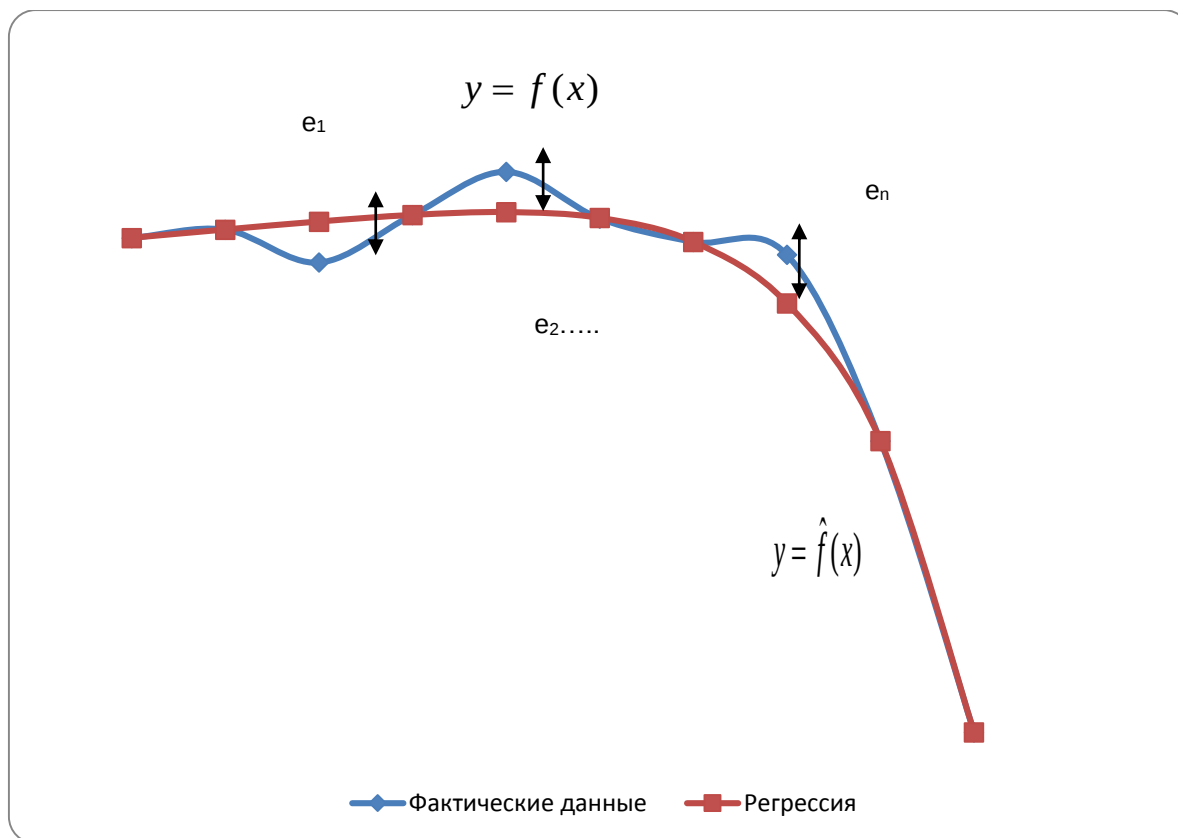


Рис. 3.1. Построение однофакторной регрессии

Регрессия и корреляция: экономический смысл и оценка параметров

Однофакторная линейная регрессия имеет вид:

$$\hat{f}(x) = a + bx + e.$$

Интерпретация коэффициентов регрессии: b – это коэффициент регрессии, показывающий насколько (как) в среднем изменится y при увеличении или уменьшении x на 1. Если $b > 0$, то наблюдается прирост y при увеличении x на единицу. Если $b < 0$, то наблюдается уменьшение y при увеличении x на единицу. Коэффициент a – свободный член уравнения регрессии (константа), обычно экономического смысла он не имеет, но иногда его интерпретируют как начальное значение y , значение y при $x=0$.

В случае если переменная x – время, рассматривается временной (динамический) ряд.

Таблица 3.1

Наиболее распространенные виды функций и их преобразование

Название	Вид функции	Вид линейной функции
Линейная	$f(x) = a + bx,$ $\hat{f}(x) = a + bx + e,$ следовательно $y = a + bx + e.$	—
Показательная	$f(x) = ab^x,$ $\hat{f}(x) = ab^x e,$ следовательно $y = ab^x e.$	$\ln y = \ln(ab^x),$ по свойству логарифмов $\ln y = \ln a + x \ln(b),$ вводим замену $\ln y = Y, \ln a = A, \ln(b) = B,$ тогда $Y = A + Bx.$
Степенная	$f(x) = ax^b,$ $\hat{f}(x) = ax^b e,$ следовательно $y = ax^b e.$	$\ln y = \ln(ax^b),$ по свойству логарифмов $\ln y = \ln a + \ln x^b$ или $\ln y = \ln a + b \ln x,$ вводим замену $\ln y = Y, \ln a = A, \ln x = X,$ тогда $Y = A + bX.$
Экспоненциальная	$f(x) = a + \frac{b}{x},$ $\hat{f}(x) = a + \frac{b}{x} + e,$ следовательно $y = a + \frac{b}{x} + e.$	$\ln y = \ln(e^{a+bx}),$ по свойству логарифмов $\ln y = a + bx,$ вводим замену $\ln y = Y,$ тогда $Y = a + bx.$
Гиперболическая	$f(x) = e^{a+bx},$ $\hat{f}(x) = e^{a+bx+e},$ следовательно $y = e^{a+bx+e}$	вводим замену $1/x = X,$ тогда $y = a + bX.$

С помощью метода наименьших квадратов можно найти коэффициенты однофакторной регрессии по формулам:

$$a = \bar{y} - b\bar{x}, \quad b = \frac{\overline{yx} - \bar{y}\bar{x}}{\overline{x^2} - \bar{x}^2}.$$

Коэффициент b называется коэффициентом регрессии. Он показывает среднее изменение результата при изменении фактора на единицу.

Например, зависимость стоимости квартиры (тыс. руб.) от общей площади квартиры (кв. м) задается уравнением регрессии $y = 145 + 6,9x$, тогда можно сделать вывод, что при увеличении общей площади квартиры на 1 кв. м ее стоимость возрастает на 6,9 тыс. руб. Коэффициент a – это значение результативной переменной при $x=0$. Экономической интерпретации он не имеет.

Оценить тесноту связи двух признаков x и y в линейной форме можно с помощью парного коэффициента корреляции:

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = \frac{\overline{yx} - \bar{y}\bar{x}}{\sigma_x \sigma_y},$$

где b – коэффициент при x в уравнении регрессии $y = a + bx$,

σ_x , σ_y – среднеквадратические отклонения по x и y .

Его интерпретация рассматривалась во второй главе.

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции r_{xy}^2 , который называется *коэффициентом детерминации*. При $r_{xy}^2 \geq 0,5$ можно говорить о высоком качестве модели, на ее основе можно проводить анализ и строить прогнозы.

Оценить тесноту связи двух признаков x и y в нелинейной форме можно с помощью *индекса корреляции*:

$$\rho_{xy} = \sqrt{1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}},$$

Значение индекса корреляции находится в пределах:

$$0 \leq \rho_{xy} \leq 1.$$

Если $\rho_{xy}=0$, можно говорить об отсутствии связи между признаками или неправильно выбранной форме связи. Если $\rho_{xy}=1$,

связь между x_i и y_i самая сильная – функциональная. Связь может быть слабая, умеренная и высокая (табл. 3.2).

Таблица 3.2

Значения индекса корреляции

Показания тесноты связи	$ 0,1 - 0,3 $	$ 0,3 - 0,5 $	$ 0,5 - 0,7 $	$ 0,7 - 0,9 $	$ 0,9 - 0,99 $
Характеристика тесноты связи	Слабая	Умеренная	Заметная	Высокая	Весьма высокая

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции ρ^2_{xy} , который называется индексом детерминации. Если $\rho^2_{xy} \geq 0,5$, можно говорить о высоком качестве модели, на ее основе можно проводить анализ и строить прогнозы.

Прогнозирование по динамическим и пространственным моделям

Как уже отмечалось качество построенной регрессионной модели считается высоким если:

1. Коэффициент детерминации (R^2) больше либо равен 0,5;
2. Критерий Фишера $F_{\text{табл}} < F_{\text{факт}}$;
3. Средняя ошибка аппроксимации ($\bar{A}$) меньше либо равна 10%.

Если все три условия выполняются, то построенная модель может быть использована для прогнозирования. Если хотя бы одно из условий не выполняется, то модель низкого качества и не может быть использована для прогнозирования.

Классический способ прогнозирования по динамическим и пространственным моделям – экстраполяция и интерполяция.

Экстраполяция – метод прогнозирования, при котором функция аппроксимирует вне рассматриваемого периода (рис. 3.2). Она предполагает нахождение уровней за пределами изучаемого ряда, т.е. продление в будущее тенденции, наблюдавшейся в прошлом.

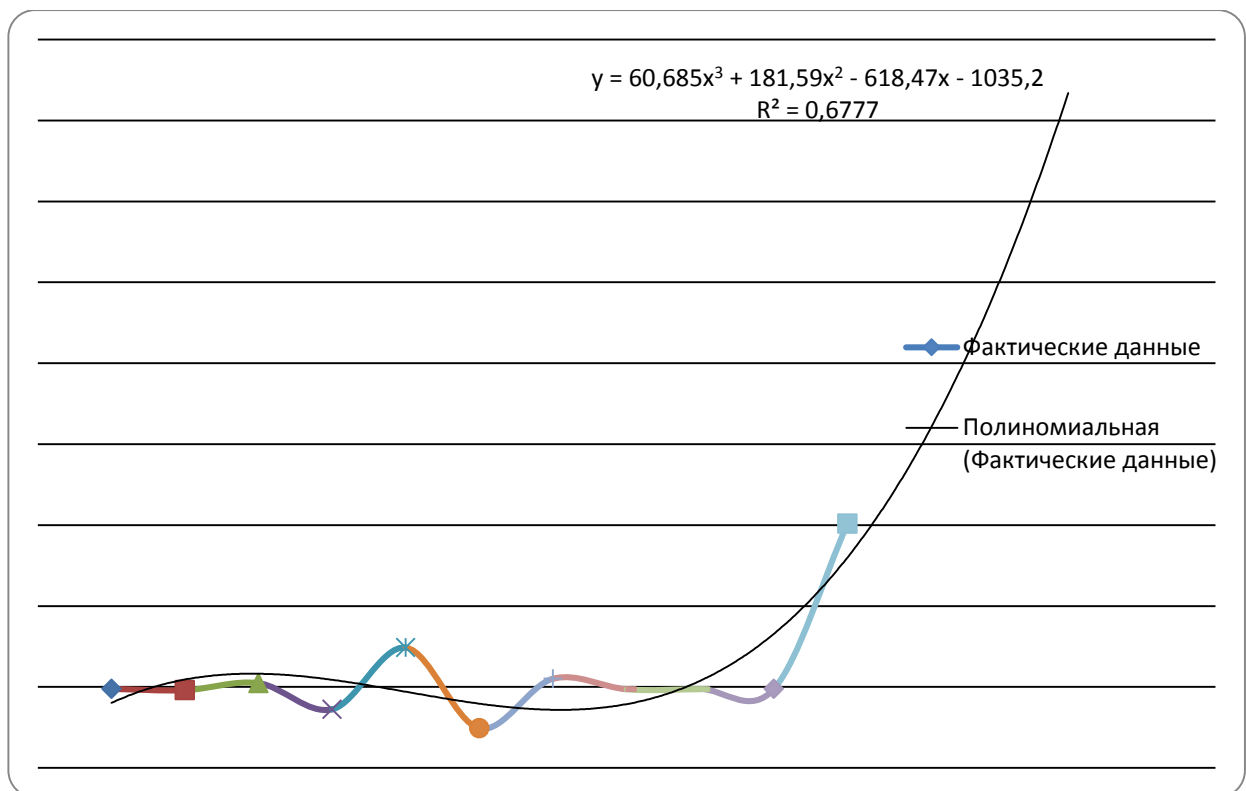


Рис. 3.2. Экстраполяция по фактическим данным полиномиальной функцией

Интерполяция – метод прогнозирования, при котором функция аппроксимирует внутри рассматриваемого периода (рис. 4.3).

Интерполяция предполагает нахождение по имеющимся данным за определенный период времени некоторых недостающих значений признака внутри этого признака.

Для этого необходимо построить регрессию.

Регрессионной моделью системы взаимосвязанных признаков является такое уравнение регрессии, которое включает основные факторы, влияющие на вариацию результативного признака, обладает высоким (не ниже 0,5) коэффициентом детерминации и коэффициентами регрессии, интерпретируемыми в соответствии с теоретическим знанием о природе связей в изучаемой системе.

Выбор наилучшего варианта регрессионной модели осуществляется сравнением их качественных характеристик. Соответственно лучшему варианту модели должны соответствовать лучшие характеристики.

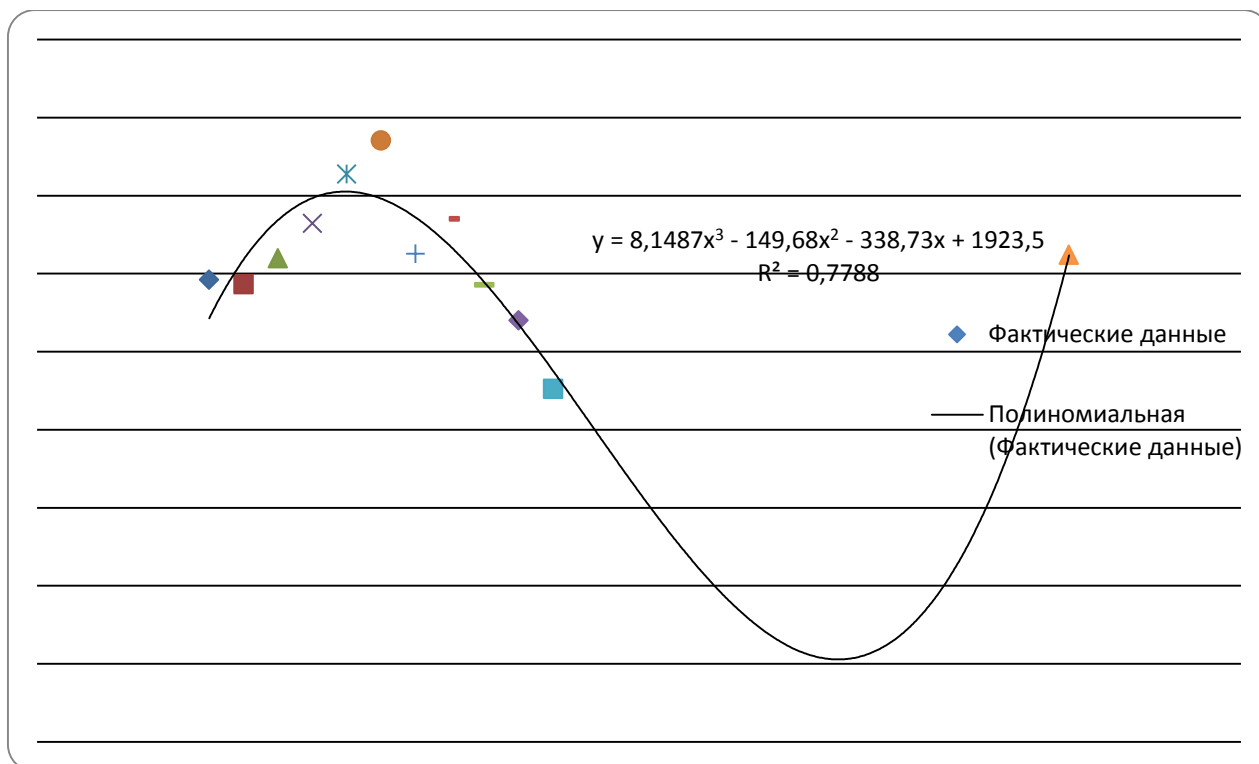


Рис. 3.3. Интерполяция по фактическим данным полиномиальной функцией на интервале (5;20)

Прогнозирование по модели однофакторной линейной регрессии (также, как и в многофакторной) предполагает оценку ожидаемых значений зависимой переменной при заданных значениях независимых переменных, входящих в уравнение регрессии. Также можно построить точечный и интервальный прогнозы.

Оценить точность прогнозов с помощью регрессии можно с помощью ошибки прогноза, показывающей относительное отклонение расчетных (регрессионных) значений от фактических. При этом оценивать ошибки прогнозов нужно при составлении прогнозов, а не после получения результатов исследований.