ПАРНАЯ НЕЛИНЕЙНАЯ КОРРЕЛЯЦИЯ

ГЛАВА 4

Содержание

§ 18. Нелинейная корреляционная зависимость

§ 19. Определение силы криволинейной связи

§ 20. Проверка адекватности модели

§ 18. Нелинейная корреляционная зависимость

Между изучаемыми признаками *X* и *Y* может существовать нелинейная корреляционная зависимость. Если линия регрессии с уравнением $\hat{y}_x = f(x)$ или $\hat{x}_y = \varphi(y)$ не является прямой, то зависимость между показателями называют нелинейной корреляционной зависимостью. Среди видов нелинейной корреляции обыкновенно выделяют: полиномиальную (в частности, параболическую), степенную (в частности, гиперболическую), экспоненциальную, гармоническую, фрактальную и другие. Рассмотрим некоторые из них, начав с параболической корреляции.

Пусть зависимость между признаками *X* и *Y* задана в виде корреляционной таблицы. Для определения типа нелинейной зависимости на координатной плоскости строят точки $M_i(x_i, y_j)$.

§ 18. Нелинейная корреляционная зависимость

Параболическая корреляция. Если точки в корреляционном поле располагаются вблизи некоторой параболы, то уравнение регрессии записывают в виде

$$\hat{y}_x = a_0 + a_1 x + a_2 x^2 .$$

§ 18. Нелинейная корреляционная зависимость

Оценки a_0, a_1, a_2 для неизвестных параметров истинного уравнения регрессии находят по методу наименьших квадратов. Если опытные данные не сгруппированы в корреляционную таблицу, то оценки находят, решая систему нормальных уравнений

$$\begin{cases} n a_0 + [x]a_1 + [x^2]a_2 = [y], \\ [x]a_0 + [x^2]a_1 + [x^3]a_2 = [xy], \\ [x^2]a_0 + [x^3]a_1 + [x^4]a_2 = [x^2 y]. \end{cases}$$

§ 18. Нелинейная корреляционная зависимость

Для сгруппированных значений признаков *X* и *Y* оценки a_0, a_1, a_2 находят, решая СЛАУ:

$$\begin{cases} n a_0 + [n_x x]a_1 + [n_x x^2]a_2 = [n_x \bar{y}_x], \\ [n_x x]a_0 + [n_x x^2]a_1 + [n_x x^3]a_2 = [n_x x \bar{y}_x], \\ [n_x x^2]a_0 + [n_x x^3]a_1 + [n_x x^4]a_2 = [n_x x^2 \bar{y}_x]. \end{cases}$$

§ 18. Нелинейная корреляционная зависимость

Гиперболическая корреляция 1. Зависимость между *X* и *Y* может быть близкой к гиперболической. В этом случае уравнение регрессии ищут в виде

$$\hat{y}_x = a_0 + \frac{a_1}{x}.$$

§ 18. Нелинейная корреляционная зависимость

Оценки a_0 и a_1 неизвестных параметров истинного уравнения регрессии находят по методу наименьших квадратов, решая систему нормальных уравнений:

$$\begin{cases} n a_0 + [1/x]a_1 = [y], \\ [1/x]a_0 + [1/x^2]a_1 = [y/x], \end{cases}$$

где

$$[1/x] = \sum_{i=1}^{n} 1/x_i\,,\ [1/x^2] = \sum_{i=1}^{n} 1/x_i^2\,,\ [y] = \sum_{i=1}^{n} y_i\,,\ [y/x] = \sum_{i=1}^{n} y_i/x_i\,.$$

§ 18. Нелинейная корреляционная зависимость

Гиперболическая корреляция 2. Если гиперболическая зависимость между признаками *X* и *Y* имеет вид

$$\hat{y}_x = \frac{1}{a_0 + a_1 x}$$

то оценки a_0 и a_1 находят, решая систему нормальных уравнений

$$\begin{cases} n a_0 + [x]a_1 = [1/y], \\ [x]a_0 + [x^2]a_1 = [x/y]. \end{cases}$$

§ 18. Нелинейная корреляционная зависимость

Если зависимость между признаками имеет экспоненциальный характер, то уравнение регрессии ищут в виде

$$\hat{y}_x = a \cdot b^x .$$

Для определения оценок *a* и *b*, входящих в уравнение регрессии,решают СЛАУ:

$$\begin{cases} [x]\lg a + [x^2]\lg b = [x\lg y], \\ n\lg a + [x]\lg b = [\lg y]. \end{cases}$$

§ 18. Нелинейная корреляционная зависимость

Если в корреляционном поле около построенных точек предполагается проведение разных по типу линий (параболы, гиперболы, экспоненты, логарифмики), то для выбора одной из них, характеризующей наилучшим образом зависимость между признаками X и Y , применяют либо метод конечных разностей, либо производят проверку необходимых условий.

§ 18. Нелинейная корреляционная зависимость. Проверка необходимых условий

Проверку необходимых условий для выбора одной из предполагаемых нелинейных зависимостей проводят, пользуясь табл. 33.

Таблица 33

Необходимые условия	Регрессия	Способ выравнивания (приведения к линейной зависимости $Y = AX + B$)
$y\left(\frac{x_1+x_n}{2}\right) = \frac{y(x_1)+y(x_n)}{2}$	линейная $y = a\,x + b$	Тождественное преобразование
$y(\sqrt{x_1 x_n}) = \sqrt{y(x_1)\,y(x_n)}$	степенная $y = a\,x^b$	$Y = \ln y,\ X = \ln x,$ $A = \ln a,\ B = b$
$y\left(\frac{x_1+x_n}{2}\right) = \sqrt{y(x_1)\,y(x_n)}$	экспоненциальная $y = a\,b^x$	$Y = \ln y,\ X = x,$ $A = \ln a,\ B = \ln b$
$y\left(\frac{2x_1 x_n}{x_1+x_n}\right) = \frac{y(x_1)+y(x_n)}{2}$	гиперболическая $y = a + \frac{b}{x}$	$Y = xy,\ X = x$
$y\left(\frac{x_1+x_n}{2}\right) = \frac{2y(x_1)\,y(x_n)}{y(x_1)+y(x_n)}$	гиперболическая $y = \frac{1}{ax+b}$	$Y = 1/y,\ X = x$
$y(\sqrt{x_1\,x_n}) = \frac{y(x_1)+y(x_n)}{2}$	логарифмическая $y = a\ln x + b$	$Y = y,\ X = \ln x$

§ 18. Нелинейная корреляционная зависимость. Проверка необходимых условий

Если выполняется одно из условий первого столбца таблицы, то выбирают в качестве предполагаемой формулы соответствующую формулу, стоящую во втором столбце таблицы рассматриваемой строки. В третьем столбце указывается способ выравнивания, то есть приведения изучаемой зависимости к линейной. Если выровненные точки (X_i , Y_i) хорошо ложатся на прямую, то указанную во втором столбце таблицы зависимость принимаем в качестве предполагаемой.

§ 18. Нелинейная корреляционная зависимость. Проверка необходимых условий

Если значения функции, вычисленные в первом столбце таблицы при выбранных значениях аргумента, отсутствуют в таблице опытных данных, то их находят линейным интерполированием по формуле

$$y(x) = y(x_1) + \frac{y(x_2) - y(x_1)}{x_2 - x_1}(x - x_1),$$

где x_1 и x_2 — два рядом стоящих значения признака X в таблице опытных данных, между которыми находится значение x , вычисленное по табл. 33 первого столбца.

§ 18. Нелинейная корреляционная зависимость. Проверка необходимых условий

Для всех предполагаемых формул по результатам первого столбца табл. 33 вычисляют отклонения Δ правой части от левой необходимого условия. Вычисленные отклонения Δ_i сравнивают и по наименьшему из них выбирают окончательно одну из формул.

§ 18. Нелинейная корреляционная зависимость. Метод конечных разностей

Пусть в корреляционном поле могут быть проведены линии, описываемые уравнениями

$$y = ax^b,$$

$$y = ae^{cx}\,(c = \ln b),$$

$$y = a + \frac{b}{x},$$

$$y = \frac{1}{ax+b},$$

$$y = a\ln x + b.$$

§ 18. Нелинейная корреляционная зависимость. Метод конечных разностей

Все эти формулы содержат по два параметра и могут быть приведены к формуле $Y = AX + B$ посредством табл. 33.

Так как все зависимости, приведенные в табл. 33, сводятся к линейной, то для обоснования выбора формулы $Y = AX + B$ вычисляют конечные разности первого порядка DX , DY и отношения DY / DX .

Аналитическим критерием выбора формулы по этому методу служит тот факт, что отношения DY / DX мало отличаются друг от друга для выбранной формулы.

§ 18. Нелинейная корреляционная зависимость. Метод конечных разностей

Если предполагаемая формула имеет вид $y = ax^2 + bx + c$, то критерием выбора этой формулы являются незначительные отклонения по модулю конечных разностей второго порядка $\Delta^2 y$ от среднего значения этих разностей $\overline{\Delta^2 y}$. Конечные разности находят, пользуясь табл. 34.

§ 18. Нелинейная корреляционная зависимость. Метод конечных разностей

Таблица 34

Таблица конечных разностей

x	y	Δy	$\Delta^2 y$
x_0	$\underline{y_0}$		
		$\underline{\Delta y_0}$	
x_1	y_1		$\underline{\Delta^2 y_0}$
		Δy_1	
x_2	y_2		$\Delta^2 y_1$
		Δy_2	
x_3	y_3		$\Delta^2 y_2$
		Δy_3	
x_4	y_4		$\Delta^2 y_3$
		Δy_4	
x_5	y_5		
…	…	…	…

§ 19. Определение силы криволинейной связи

Для определения тесноты связи между признаками X и Y при нелинейной корреляции используют корреляционные отношения и индекс корреляции. Корреляционным отношением η_{yx} называется величина, определяемая равенством:

$$\eta_{yx} = \sqrt{\frac{\left(\sum n_x (\bar{y}_{xi} - \bar{y})^2\right)/n}{\left(\sum n_y (y - \bar{y})^2\right)/n}},$$

где n — объем выборки, n_x — частота значения x признака X , n_y — частота значения y признака Y , $\bar{y}$ — общая средняя признака Y , $\bar{y}_{xi}$ — условная средняя признака Y .

§ 19. Определение силы криволинейной связи

Аналогично определяется выборочное корреляционное отношение η_{xy}:

$$\eta_{xy} = \sqrt{\frac{\left(\sum n_y (\bar{x}_{yi} - \bar{x})^2\right)/n}{\left(\sum n_x (x - \bar{x})^2\right)/n}}\,.$$

где n — объем выборки, n_x — частота значения x признака X , n_y — частота значения y признака Y , $\bar{x}$ — общая средняя признака X , $\bar{x}_{yi}$ — условная средняя признака X .

§ 19. Определение силы криволинейной связи

Корреляционные отношения обладают следующими свойствами (сформулируем свойства для η_{yx}, так как для η_{xy} они аналогичны):

1. Корреляционное отношение заключено между 0 и 1, т.е. $0 \leq \eta_{yx} \leq 1$.
2. Корреляционная связь между признаками *X* и *Y* отсутствует тогда и только тогда, когда $\eta_{yx} = 0$.
3. Если $\eta_{yx} = 1$, то между признаками *X* и *Y* существует обычная функциональная связь.
4. Чем ближе значение η_{yx} к 1, тем сильнее корреляционная связь между признаками *X* и *Y* , а чем ближе η_{yx} к 0, тем слабее эта зависимость.
5. Если $\eta_{yx} = |r|$ регрессия *y* на *x* является линейной.
6. Коэффициент линейной корреляции *r* не превосходит по модулю, то есть $| r | \leq \eta_{yx}$

§ 19. Определение силы криволинейной связи

По коэффициенту корреляции r можно судить о наличии и тесноте линейной корреляционной связи между признаками X и Y.

По корреляционным отношениям можно судить только о наличии и силе корреляционной связи между признаками X и Y, но не о форме связи, которая устанавливается из *г е о м е т р и ч е с к и х* соображений.

§ 19. Определение силы криволинейной связи

Теснота связи между признаками *X* и *Y* при любой форме корреляции может быть измерена также с помощью индекса корреляции *i*. Если опытные данные не сгруппированы в корреляционную таблицу, то индекс корреляции находят по формуле:

$$i = \sqrt{1 - \frac{S_{yx}^2}{S_y^2}},$$

$$S_{yx}^2 = \left(\sum (y_j - \hat{y}_{x_j})^2\right) / n$$

— средний квадрат отклонений фактических значений y от значений y , вычисленных по уравнению регрессии;

$$S_y^2 = \left(\sum (y_j - \bar{y})^2\right) / n$$

— средний квадрат отклонений фактических значений y от их средней арифметической.

§ 19. Определение силы криволинейной связи

Если опытные данные сгруппированы в корреляционную таблицу, то индекс корреляции находят по формуле

$$i = \sqrt{1 - \frac{Q_e}{Q}},$$

или

$$i = \sqrt{\frac{\sigma_y^2}{S_y^2}},$$

где

$$Q_e = Q - Q_R, \; Q = \sum (y_j - \bar{y})^2, \; Q_R = \sum (\bar{y}_{x_i} - \bar{y})^2,$$

$$S_y^2 = (\sum (y_j - \bar{y})^2 n_j)/n, \; \sigma_y^2 = \sum (\bar{y}_{x_i} - \bar{y})^2 \cdot n_i / n.$$

§ 19. Определение силы криволинейной связи

Индекс корреляции по величине изменяется от 0 до 1. По индексу корреляции можно определять, как правило, тесноту связи между признаками X и Y, но не форму криволинейной связи.

§ 20. Проверка адекватности модели

Проверить адекватность модели — значит установить, соответствует ли построенное уравнение регрессии опытным данным и достаточно ли включенных в уравнение факторных признаков *Xi* для описания результативного признака. Оценка значимости уравнения регрессии производится на основе дисперсионного анализа. В случае нелинейной парной корреляции находят статистику $F_{\text{Н}}$ по формуле:

$$F_{\text{н}} = \frac{Q_R(n-2)}{Q_e}$$

или

$$F_{\text{н}} = \frac{i^2(n-2)}{1-i^2}.$$

§ 20. Проверка адекватности модели

Если оценке подвергаются не два, а k параметров уравнения регрессии, то находят статистику

$$F_{\text{н}} = \frac{Q_R(n-k)}{Q_e(k-1)}.$$

Затем при заданном уровне значимости α и числах степеней свободы k_1=1, k_2= n - 2 по таблице критических точек распределения Фишера — Снедекора находят (приложение 7) $F_T = F_{\alpha:k1:k2}$. Если $F_{\text{Н}} > F_T$, то уравнение регрессии согласуется с опытными данными, в противном случае — нет.

Спасибо за внимание!