

⁴⁸DOI: 10.58168/E-SESTTI2025_211-216

УДК 625.7/.8

ИССЛЕДОВАНИЕ И РАЗРАБОТКА МОДЕЛЕЙ ПРОГНОЗА ПРОДАЖ НА ОСНОВЕ МЕТОДОВ КОРРЕЛЯЦИИ И РЕГРЕССИИ

RESEARCH AND DEVELOPMENT OF SALES FORECAST MODELS BASED ON CORRELATION AND REGRESSION

Анциферова В.И. кандидат технических наук, доцент, ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова», Воронеж, Российская Федерация

Antsiferova V.I. Candidate of Technical Sciences, Associate Professor, Voronezh State University of Forestry and Technologies named after G.F. Morozov, Voronezh, Russian Federation.

Петров Б.Я., магистр группы ИС4-231-ОМ, факультета компьютерных наук и технологий, ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова», Воронеж, Российская Федерация

Petrov B.Y., Master of IS4-231-OM group, Faculty of Computer Science and Technology, Voronezh State University of Forestry and Technologies named after G.F. Morozov, Voronezh, Russian Federation.

Аннотация: В данной работе рассматривается теоретическое изложение уравнения регрессии и его применение в различных областях исследования. Уравнение регрессии представляет собой математическую модель, используемую для описания зависимости между одной зависимой переменной и одной или несколькими независимыми переменными.

Abstract: This work discusses the theoretical exposition of the regression equation and its application in various fields of research. The regression equation is a mathematical model used to describe the relationship between one dependent variable and one or more independent variables.

Ключевые слова: Планирование, экономико-математические подходы, регрессионный анализ, критерий корреляции Пирсона.

Keywords: Planning, economic and mathematical approaches, regression analysis, Pearson correlation criterion.

Критерий корреляции Пирсона – это метод параметрической статистики, дающий возможность определить присутствие или отсутствие линейной связи между двумя количественными показателями, оценить ее статистическую значимость и её тесноту. Другими словами, критерий корреляции Пирсона, позволяет определить, имеется ли линейная связь между изменениями значений двух представленных переменных.

Критерий корреляции Пирсона может выявить тесноту корреляционной связи различных показателей, которые будут измерены по шкале. С помощью большого количества расчетов будут определены важные аспекты, один из таких аспектов – выявление связи.

Условиями и ограничениями, для применения критерия Хи-квадрата Пирсона являются:

1. Сопоставляемые показатели, измеренные только в количественной шкале.

⁴⁸ Анциферова В.И., Петров Б.Я., 2025

2. С помощью критерия корреляции Пирсона появляется возможность в определении наличия силы линейной связи. С помощью регрессионного анализа, можно с легкостью найти наличие зависимости между двумя переменными и её определить.

3. Сумма исследуемых величин всегда должна быть равна двум. В различных ситуациях можно сделать анализ взаимосвязи четырех параметров, для чего необходим метод факторного анализа.

4. Критерий корреляции Пирсона является параметрическим, и его применение требует, чтобы сравниваемые переменные имели нормальное распределение. Если необходимо провести корреляционный анализ переменных с ненормальным распределением, включая показатели, измеренные по порядковой шкале, следует использовать коэффициент ранговой корреляции Спирмена.

5. Необходимо различать и разграничивать понятия зависимости и корреляции. Зависимость величин влечет наличие корреляционной связи между ними, но не наоборот.

Самым известным методом анализа связи между количественными переменными, это – метод регрессионного анализа. В самом известном эксперименте мы видим значения $t+1$ переменных $-Y, X_1, \dots, X_t$.

Регрессионный анализ изучает связь между зависимой переменной Y и независимыми переменными $X_1, X_2, \dots, X_t$. Эта связь формулируется с использованием определенной математической модели.

Чтобы получить матрицу экспериментальных данных нужно воспользоваться одним из 2 способов. При использовании первого способа значения независимых переменных $X_1, X_2, \dots, X_t$ будут выбираться, и устанавливаться погрешность их измерения, в каждом опыте они будут измеряться с ошибками. Такой эксперимент является активным, при данном подходе случайные величины исключаются.

С помощью второго способа - все переменные $X_1, X_2, \dots, X_t$ наблюдаются и все данные случайны, другими словами матрица в данной ситуации случайна, выборка значений многомерной величины $(X_1, X_2, \dots, X_t)$.

Второй способ проводят с помощью корреляционного анализа, в таком случае статистический вывод, о возможности линейной зависимости между указанными переменными, зависит от параметров корреляции.

Самым известным и часто используемым является метод регрессионного анализа, метод наименьших квадратов. Благодаря этому методу можно получить математическое описание объекта.

По имеющимся значениям $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

Необходимо построить функцию $f(x)$.

Уравнение регрессионной зависимости выходной величины от переменных факторов выполняется в следующей последовательности:

- 1) выбираем тип уравнения;
- 2) вычисляем коэффициенты регрессии;
- 3) проверяем адекватность модели.

Рассмотрим данную регрессионную модель. Предположим, что случайная величина X зависит от T , принимая значения $t_1, t_2, \dots, t_n$, которые не являются случайными.

Зависимость среднего значения случайной величины X от T будет отображаться значением, называемым математическим ожиданием и определяемым по формуле

$$M(X|T = t_i) = f(t), \quad (1)$$

где $X = f(t)$ – уравнение регрессии,

$f(t)$ представляет собою кривую регрессии X на T .

Разница, между математическим ожиданием и полученной в данном эксперименте величиной X , вычисляется по формуле

$$\varepsilon_i = X_i - M(X|T = t_i) = X_i - f(t_i), \quad (2)$$

и есть ошибка наблюдения.

Ошибки наблюдений, должны удовлетворять условия Гаусса-Маркова:

1) математическое ожидание $M(\varepsilon_i) = 0$, при всех значениях i ;

2) дисперсия $D(\varepsilon_i) = \sigma^2 < +\infty$ при всех значениях i ;

3) величины независимы $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$, при всех значениях i, j .

Набор $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$ является элементарными событиями одной и той же случайной величины, вследствие того, что при всех значениях ε_i $M(\varepsilon_i)$ и $D(\varepsilon_i)$ одинаковы.

Дисперсию величины $\mathcal{E}$, определим по формуле

$$D[\mathcal{E}] = M[\mathcal{E}^2] - (M[\mathcal{E}])^2 = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 \quad (3)$$

или

$$D[\mathcal{E}] = \frac{1}{n} \sum_{i=1}^n (X_i - f(t_i))^2$$

Допустив, что функция $f(t)$ определяется параметрами $\vec{c} = \{c_1, c_2, \dots, c_m\}$ или вектором $\vec{c} = \{c_1, c_2, \dots, c_m\}$ имеем $f(t, \vec{c})$ и тогда дисперсия ошибки от неизвестных параметров равна

$$D[\mathcal{E}] = \frac{1}{n} \sum_{i=1}^n (X_i - f(t_i, \vec{c}))^2 \quad (4)$$

Дисперсия ошибок $D[\mathcal{E}] \subset (0, +\infty)$ и имеет минимум на интервале $(0, \infty)$, который определим из выражения

$$\frac{\partial D[\mathcal{E}]}{\partial \vec{c}} = \frac{\partial D[\mathcal{E}]}{\partial c_0} \cdot \vec{e}_0 + \dots + \frac{\partial D[\mathcal{E}]}{\partial c_m} \cdot \vec{e}_m = 0 \quad (5)$$

где $\{\vec{e}_k\}_{k=0}^m = \{\vec{e}_0, \dots, \vec{e}_m\}$ есть базис из линейно-независимых функций

$\{\vec{e}_k\}_{k=0}^m = \{\phi_k\}_{k=0}^m = \{\phi_0(x), \dots, \phi_m(x)\}$ при условии $m < n$ аппроксимирующая

функция есть разложение по базису: $f(x, \vec{c}) = \sum_{k=0}^m c_k \cdot \phi_k(x)$.

Дисперсию можно записать в виде:

$$D[\varepsilon] = \frac{1}{n} \sum_{i=1}^n (y_i - \sum_{k=0}^m c_k \cdot \phi_k(x_i))^2 \quad (6)$$

Для нахождения коэффициентов регрессии продифференцируем по каждому c_j из $\{c_k\}_{k=0}^m$, где $j = 0, 1, 2, \dots, m$.

$$\begin{cases} \frac{\partial D}{\partial c_0} = 0 \\ \frac{\partial D}{\partial c_1} = 0 \\ \dots \\ \frac{\partial D}{\partial c_m} = 0 \end{cases} \quad (7)$$

$$\frac{\partial D}{\partial c_j} = \frac{\partial}{\partial c_j} \sum_{i=1}^n (y_i - \sum_{k=0}^m c_k \cdot \phi_k(x_i))^2 = -2 \cdot \sum_{i=1}^n (y_i - \sum_{k=0}^m c_k \phi_k(x_i)) \cdot \sum_{k=0}^m \frac{\partial c_k}{\partial c_j} \phi_k(x_i) = 0 \quad (8)$$

$$\frac{\partial c_k}{\partial c_j} = \begin{cases} 1, j = k \\ 0, j \neq k \end{cases} \equiv \delta_{jk} \quad (9)$$

Учитывая, что $\sum_{k=0}^m \delta_{jk} \phi_k(x_i) = \phi_j(x_i)$, имеем

$$\sum_{i=1}^n (y_i - \sum_{k=0}^m c_k \phi_k(x_i)) \cdot \phi_j(x_i) = 0, \quad (10)$$

где $j = 0, 1, 2, \dots, m$.

Полученное выражение представим в виде системы уравнений:

$$\begin{cases} \sum_{i=1}^n (y_i - \sum_{k=0}^m c_k \phi_k(x_i)) \cdot \phi_0(x_i) = 0 \\ \sum_{i=1}^n (y_i - \sum_{k=0}^m c_k \phi_k(x_i)) \cdot \phi_1(x_i) = 0 \\ \dots \\ \sum_{i=1}^n (y_i - \sum_{k=0}^m c_k \phi_k(x_i)) \cdot \phi_m(x_i) = 0 \end{cases} \quad (11)$$

В матричном виде система уравнений будет иметь вид: $A \cdot \vec{c} = \vec{b}$,

$$A = \begin{pmatrix} (\phi_0, \phi_0)(\phi_1, \phi_0) \dots (\phi_m, \phi_0) \\ (\phi_0, \phi_1)(\phi_1, \phi_1) \dots (\phi_m, \phi_1) \\ \dots \\ (\phi_0, \phi_m)(\phi_1, \phi_m) \dots (\phi_m, \phi_m) \end{pmatrix}, \quad \vec{c} = \begin{pmatrix} c_0 \\ c_1 \\ \dots \\ c_m \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} (y, \phi_0) \\ (y, \phi_1) \\ \dots \\ (y, \phi_m) \end{pmatrix} \quad (12)$$

где

Решая систему, получаем вектор коэффициентов аппроксимирующей функции:

$$\vec{c} = \{c_0, c_1, \dots, c_m\} = A^{-1} \cdot b = \begin{pmatrix} (\varphi_0, \varphi_0)(\varphi_1, \varphi_0) \dots (\varphi_m, \varphi_0) \\ (\varphi_0, \varphi_1)(\varphi_1, \varphi_1) \dots (\varphi_m, \varphi_1) \\ \dots \\ (\varphi_0, \varphi_m)(\varphi_1, \varphi_m) \dots (\varphi_m, \varphi_m) \end{pmatrix}^{-1} \begin{pmatrix} (y, \varphi_0) \\ (y, \varphi_1) \\ \dots \\ (y, \varphi_m) \end{pmatrix} \quad (13)$$

Аппроксимирующая функция в результате будет иметь вид:

$$f(x) = c_0 \varphi_0(x) + c_1 \varphi_1(x) + \dots + c_m \varphi_m(x) \quad (14)$$

В случае выбора степенной функции аппроксимирующую записывают в виде:

$$f(x, \vec{c}) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \dots + c_m x^m = \sum_{k=0}^m c_k x^k \quad (15)$$

Система $A \cdot \vec{c} = \vec{b}$ примет вид:

$$A = \begin{pmatrix} \sum x_1 & \sum x_i & \dots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \dots & \sum x_i^{m+1} \\ \dots & \dots & \dots & \dots \\ \sum x_i^m & \sum x_i^{m+1} & \dots & \sum x_i^{2m} \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} \sum y_i \\ \sum y_i x_i \\ \dots \\ \sum y_i x_i^m \end{pmatrix} \quad (16)$$

$$\vec{c} = \begin{pmatrix} c_0 \\ c_1 \\ \dots \\ c_m \end{pmatrix} = A^{-1} b = \begin{pmatrix} \sum x_1 & \sum x_i & \dots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \dots & \sum x_i^{m+1} \\ \dots & \dots & \dots & \dots \\ \sum x_i^m & \sum x_i^{m+1} & \dots & \sum x_i^{2m} \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i \\ \sum y_i x_i \\ \dots \\ \sum y_i x_i^m \end{pmatrix} \quad (17)$$

где

$$f(x) = \begin{pmatrix} c_0 \\ c_1 \\ \dots \\ c_m \end{pmatrix} \{1, x, \dots, x^m\} = f(x, \vec{c}) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \dots + c_m x^m \quad (18)$$

Модель такого вида применять нельзя, если истинная зависимость отклика $y = f(x)$ от некоторого фактора:

- 1) Будет иметь более одного экстремума;
- 2) Имеет точку перегиба;
- 3) при некоторых значениях действующих факторов функция изменяется скачком.

При возникновении первых двух случаев рекомендуется выбирать многочлены более высокого порядка.

Список литературы

1. Антамошкина, О.И. Имитационные модели прогноза продаж / О.И. Антамошкина, Ю.В. Булгаков, О.В. Зинина // Вестник КРАСГАУ. – 2011. – №2(53). – С.28-36.
2. Бахтеева, Р.Х. Статистический анализ себестоимости от продаж и оказанных услуг / Р.Х. Бахтеева // Синергия наук. – 2018. – № 20. – С. 203-210.
3. Глухова, Л.В. Оптимизация планирования производства на основе моделирования прогноза объемов продаж / Л.В. Глухова, И.А. Маштаков // ВестникВолжскогоуниверситетаим.В.Н. Татищева. – 2012. – №1(25). – С. 110-117.
4. Юров, А.Н. Проектирование автоматизированной системы производственных планировок / А.Н. Юров // Моделирование систем и процессов. – 2019. – Т. 12, № 1. – С. 87-93.
5. Построение интеллектуальных систем управления информационными процессами в условиях неопределенности / Ю.Ю. Громов, В.Е. Дидрих, И.В. Дидрих, А.Ю. Гречушкина // Моделирование систем и процессов. – 2018. – Т. 11, № 1. – С. 10-14.
6. Оксюта, О.В. Разработка математической модели оптимального функционирования транспортно-логистического комплекса / О.В. Оксюта, В.А. Коротких // Моделирование систем и процессов. – 2017. – Т. 10, № 3. – С. 55-66.

References

1. Antamoshkina, O.I. Simulation models of sales forecast / O.I. Antamoshkina, Yu.V. Bulgakov, O.V. Zinina // Vestnik KRASGAU. - 2011. - №2(53). - P. 28-36.
2. Bakhteeva, R.Kh. Statistical analysis of the cost of sales and rendered services / R.H. Bakhteeva // Synergy of Sciences. - 2018. - № 20. - P. 203-210.
3. Glukhova, L.V. Optimization of production planning on the basis of sales volume forecast modeling / L.V. Glukhova, I.A. Mashtakov // VestnikVolzhskyUniversityim.V.N. Tatishchev. - 2012. - No.1(25). - P. 110-117.
4. Yurov, A.N. Designing an automated system of production layouts / A.N. Yurov // Modeling of systems and processes. - 2019. - T. 12, № 1. - P. 87-93.
5. Building intelligent control systems for information processes in uncertainty conditions / Yu.Y. Gromov, V.E. Diedrich, I.V. Diedrich, A.Y. Grechushkina // Modeling of systems and processes. 2018. Vol. 11, No. 1. - P. 10-14
6. Oksyuta, O.V. Development of mathematical model of optimal functioning of transport and logistics complex / O.V. Oksyuta, V.A. Korotkikh // Modeling of systems and processes. -2017. -T. 10, № 3. - P. 55-66.