Kirkuk Journal of Science Vol. 20, Iss. 3, p. 78-86, 2025 doi:10.32894/kujss.2025.158402.1199

Predictive Analysis for Market Sales Using Polynomial Regression Models

Hemn Hussein Yaseen[1*], Paree Khan Abdulla Omer [1]

[1] Department of Statistics and Informatics, College of Administration and Economics, University of Salahaddin, Erbil, Iraq.
*Corresponding author : hemnblbas95@gmail.com

Article Information

Article Type:
Research Article

Keywords:
Multiple linear regression model; Polynomial regression model; least squares error

History:
Received: 9 April 2025
Reviced: 3 August 2025
Accepted: 4 August 2025
Published: 30 September 2025

Citation: Hemn Hussein Yaseen and Paree Khan Abdulla Omer, Predictive Analysis for Market Sales Using Polynomial Regression Models., Kirkuk Journal of Science, 20(3), p.78-86, 2025, https://doi.org/10.32894/kujss.2025.158402.1199

Abstract

In this paper, we analyze sales data for two selected items from a supermarket and apply polynomial regression models to uncover insights into their sales trends. By fitting a polynomial curve to the data, we aim to understand how such pricing affects the number of sales of these items. This analysis compares the effectiveness of various regression models in predicting the price of products in different contexts. The linear regression model is the most effective for predicting the price of tomato paste, as it gives the data a better fit, with almost identical MSE and SEE, indicating similar prediction accuracy and errors. The linear regression model is the most effective for predicting the dependent variable, which is the number of sales (tomato paste) items, highlighting that the price of the item (independent variable) is statistically significant and provides a clear relationship with the dependent variable. In the context of sales, the number of oil items as the dependent variable prediction, the quadratic model is the most effective for explaining the price of oil items with both independent variables (price of power one and price of power two) being statistically significant, suggesting that the quadratic model best captures the connection between the two variables.

1. Introduction:

An assessment of consumer behavior along with sales patterns allows retail operations to boost their inventory controls and pricing methods and achieve higher profits. Retail stores, with their wide range of merchandise, contain substantial information that helps identify vital sales patterns for individual products. The analysis of these patterns requires polynomial distribution as a statistical tool to create forensic models between two variables for making sales-based predictions about upcoming trends. The distribution called the polynomial distribution serves statistical needs by defining complicated variable interactions without basic linear regression patterns. A polynomial function combines various terms that combine exponents of variables with coefficients. The highest exponent value in a polynomial model establishes its range of adaptability for data adjustment. A second-degree polynomial creates parabolic curves, but third-degree polynomials introduce more advanced cubic relationships. Sales data analysis utilizes polynomials as tools to predict item sales development based on various time-dependent influences. Evaluating data with polynomials gives us the ability to identify rising, falling, and periodic behavior in sales patterns, which traditional linear regression models lack. Polynomial functions have existed across multiple studies for a long period, especially in environmental sciences, where they serve to describe intricate relationships between different variables. Different models characterized by various equation degrees provide flexibility for modeling nonlinear data dynamics [1]. The polynomial regression tool offers strong abilities to forecast waste production along with emissions and evaluate significant environmental factors required in waste management studies. This method produces alternative connections between complex mathematical curves that were derived from past data records [2]. Past research indicates that polynomial models surpass linear models because they better demonstrate complex connections between waste production and GDP per capita changes [3]. Because they monitor time-dependent changes linked to income levels, recycling behavior, and consumption patterns, third-degree polynomials demonstrated effective performance in trash generation trend simulation. This was because they tracked these changes [4]. Using polynomial regression modeling allows accurate insights into future waste effects and material shifts, which represents an effective method for understanding complex waste systems. Researchers commonly implement polynomials for environmental models to analyze complex systems that occur during waste management operations. Polynomial regression serves as an effective mathematical tool to analyze multifaceted relationships between temperature, moisture content, and airflow rates, which operate during mechanical-biological treatment and bio-drying processes [5].

Researchers have established that polynomial regression models efficiently optimize bio drying systems by identifying relationships between temperature elevation and moisture reduction throughout time, as well as employing higher-order polynomial functions to forecast net heating value in processed waste products [6]. The research makes use of third-degree polynomial models as a simulation tool to investigate the impacts of temperature and airflow on the efficiency of waste drying for the goal of industrial optimization [4].

The modeling relationship between sales data and time-based independent variables, including pricing and promotional activities, occurs through Polynomial Regression Models. The assumption of straight-line relationships from simple linear regression does not apply to polynomial regression because it enables more precise modeling of actual market behavior through curved relationships. Research examines polynomial regression models to understand sales influencers and patterns while predicting market sales for the future. The methodology incorporates market data with past sales statistics to develop precise sales predictions, which help anticipate customer needs and support operational decision-making. The research examines the ability of polynomial regression to estimate market sales and investigates prediction accuracy levels for business market behavior understanding. This paper evaluates the implementation process of sales prediction using polynomial regression models by identifying both possible advantages and implementation challenges[7].

Analyzing consumer behavior alongside sales patterns enables retail operations to improve inventory control and pricing strategies, ultimately increasing profitability. Retail stores, with their diverse merchandise offerings, hold a wealth of data that can be used to uncover important sales trends for individual products [8]. The analysis of these patterns often requires the use of polynomial distribution as a statistical tool to develop forensic models that explore relationships between variables and generate sales-based predictions about emerging trends [9]. Polynomial distribution supports such analyses by capturing complex interactions between variables that cannot be adequately modeled using basic linear regression techniques [10]. A polynomial function consists of multiple terms, each combining variable exponents with corresponding coefficients [11]. The highest exponent, known as the degree of the polynomial, determines its flexibility and ability to model complex data patterns [8]. For instance, second-degree polynomials produce parabolic curves, while third-degree polynomials capture more intricate cubic relationships [12]. In sales data analysis, polynomial models are valuable tools for predicting product performance influenced by time-dependent variables [13]. By evaluating data through polynomial functions, analysts can detect upward or downward trends, as well as cyclical behaviors in sales patterns that traditional linear regression models often fail to capture [14]. The main objective of this paper is to analyze sales data for two selected supermarket items using polynomial regression models to gain insights into their sales trends. By fitting polynomial curves to the data, the study explores how pricing influences the quantity of items sold. Additionally, the analysis compares the effectiveness of different regression models in capturing and predicting sales behavior across various pricing scenarios.

2. Materials and Methods:

2.1 The Multiple Linear Regression Model:

Regression applications that involve many predictor variables are referred to as a multiple linear regression model. Polynomial regression is a technique used in multiple regression. A response variable is regressed on the powers of the independent variables in a polynomial regression model. The basic multiple regression model of a dependent (response) variable Y on a set of k independent (predictor) variables $x_1, x_2, \ldots\ldots, x_k$ can be expressed as:

$$\begin{cases} y_1 = \beta_0 + \beta_1 X_{11} + \ldots + \beta_k X_1 k + e_1 \\ y_2 = \beta_0 + \beta_1 X_{21} + \ldots + \beta_k X_1 k + e_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n = \beta_0 + \beta_1 X_{n1} + \ldots + \beta_k X_{nk} + e_n) \end{cases} \quad (1)$$

$$y_n = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + e_n \text{ for } i = 1, 2, \ldots, n \quad (2)$$

In a model of multiple linear regression, Y_i represents the response variable's value for the i^{th} observation. X_{ij} indicates the value of the j_{th} predictor variable for the observation i^{th}. β_0 stands for the intercept of the regression plane (considering multiple dimensions), and eachβ_j (for $j = 1, 2, ..., k$) denotes the slope of the regression plane regarding the variable $j_{th} X_j$. The terme_i is the random error for the case i_{th}. In this model, we have (n) observations and k predictors, with the condition that $(nk > 1)$. The assumptions underlying the multiple regression model are similar to those of simple linear regression [15].

2.1.1 Assumptions of the Multiple Regression Model:

For each observation, the error terms follow a normal distribution with a zero mean and a stastandard deviation σ.e error terms are independent of each other, meaning the error for one observation does not affect the error for any other observation. Furthermore, the errors are uncorrelated, meaning that there is no relationship between the errors of different observations. That is $ei\ N(0,\sigma^2)$ for all i when $i = 1,2,...,n$, errors are independently for each other. While the variables Xj are random variables in correlation analysis, they are fixed quantities in regression analysis. Regardless of the error term, Xj is independent. Assuming that Xj are fixed values means that we have realizations of k variables Xj and that the error term is the only source of randomness in Y. When using matrix notation, we can rewrite model (1) as:

$$Y = X\beta + e \tag{3}$$

The response vector Y and the error vector e are both column vectors with a length of n, where n is the number of observations. The parameter vector β is a column vector of length $k+1$, representing the coefficients of the regression model (including the intercept). The design matrix X is an $n\times(k+1)$ matrix, where the first column consists of all ones (representing the intercept), the second column contains the observed values of X1 , and so on for each predictor variable. The goal is to estimate the unknown values of β (the regression coefficients) and e (the errors) [15].

2.1.2 Least Squared Eror Approach in Matrix form:

We use the least squares method to estimate the regression parameters. The process employed in simple linear regression is extended in this way. The sum of the squared errors is first determined, and then a collection of estimators that minimize the sum is identified. Equation (3) is used to determine the errors [16].

$$e = Y - X\beta \tag{4}$$

to find estimator $\hat{\beta}$ we have to minimize the errors sum of squares as:
Consider the fractional system involving the Caputo fractional derivative of the form:

$$e^T e + (Y - X\beta)^T (Y - X\beta) \tag{5}$$

where (T) denotes the transpose of the matrix. Here e^T e is scalar. We can take the first delivery of this object function with respect to the vector β. Making these equal to 0 (a vector of zeros) we obtain normal equations:

$$X^T X\hat{\beta} = X^T Y \tag{6}$$

Multiply the inverse matrix of $(X^T X)^{-1}$ on both sides in equation (6), and we will have the least squared estimator for the multiple linear regression model in matrix form [8]

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{7}$$

Vector $\hat{\beta}$ is an unbiased estimator of β. The fitted (predicted) values for the mean of Y (let us call them Y), are computed by:

$$\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y = HY \tag{8}$$

Where $H = X(X^T X)^{-1} X^T$. We call this the hat matrix because it turns Y into $\hat{Y}$. Matrix H is symmetric, i.e. $H = H^T$ and idempotent, i.e. $H^2 = H$ The fitted values for the error terms e_i are residuals $\hat{e}_i, i = 1,2...,n$, that are computed by

$$\hat{e} = Y - \hat{Y} = Y - HY = (1-H)Y \tag{9}$$

where I is an identity matrix. The sum of squares of the residuals $SSE = \hat{e}^T \hat{e}T$ has the X^2 distribution with $df_E = n-(k+1)$ degrees of freedom and is independent of$\hat{\beta}$[17].

2.2 Polynomial Regression Model and Evaluation of its Accuracy:

A special case of multiple regression is polynomial regression, with only one response variable X. A one-variable polynomial regression model can be expressed as:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \cdots + \beta_k x_i^k + e_i, \text{ for } i = 1,2...,n \tag{10}$$

where k is the degree of the polynomial. The degree of the polynomial is the order of the model. Effectively, this is the same as having a multiple model with $X_1 = X, X_2 = X^2, X_3 = X^3$, etc. The mean squared error MSE is an unbiased estimator of the variance,$\sigma^2$of the random error term and is defined in equation:

$$MSE = \frac{SSE}{df_E} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-(k+1)} \tag{11}$$

where y_i are observed values and $\hat{y}_i$ are the fitted values of the dependent variable Y for the i^{th} case. Since the mean squared error is the average squared error, where averaging is done by dividing by the degrees of freedom, MSE is a measure of how well the regression fits the data. The square root of MSE is an estimator of the standard deviation σ of the random error term. The root mean squared error $RMSE = \sqrt{MSE}$ is not an unbiased estimator of σ, but it is still a good estimator. MSE and RMSE are measures of the size of the errors in regression and do not indicate the explained component of the regression fit [18].

2.3 Models and Formulas used in the Applied Part:

There are different orders of models of regression analysis, see Figure 1.

2.3.1 Linear Regression:

Linear regression models a straight-line relationship between the dependent variable (y) and the independent variable (x):

$$y_i = \beta_0 + \beta_1 x_i + e_i \tag{12}$$

β_0: Intercept (constant term).
β_1: Slope (rate of change in y for a unit change in x).
e_i: Error term (captures deviations from predicted values).

2.3.2 Quadratic Regression:

Extends linear regression by adding a squared term to model parabolic relationships:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i \tag{13}$$

2.3.3 Cubic Regression:

Incorporates$x^3$to capture relationships with multiple inflection points or S-shaped patterns:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + e_i \tag{14}$$

3. Data Collection and Analysis:

The data for this paper investigates the price sales of two items, paste tomato and oil, and their effects on the number of items sales in the specific market known as Golden Company in Erbil, Iraq, for the production and marketing of tomato paste & oil that has been experienced in this field for many years. it showed us 312 observations had been used for this search: where 156 for oil and 156 for tomatoes paste, then analyzed the data by polynomial regression model in different power.

3.1 Discussion of Results of the Sale of Tomato Paste:

Among the three models, as presented in Table 1 and Figure 2, the linear regression model is the best, with the least mean square error-presented at 3635.405 and a standard error of the estimate of 60.294. This model strikes a balance between simplicity and performance. Finally, the quadratic and cubic models add unnecessary complexity without improving predictive accuracy. Both models have the same metrics, having an MSE of 3643.968 and a standard error of 60.365, showing again how neither of them outperforms the simpler Linear model.

Table 2 reveals that a cubic regression model best captures the relation between the price of X and the dependent variable, as evidenced by the significant cubic term $(B = -81.906, p = 0.040)$. While the linear model shows a significant negative relationship $(B = -0.107, p = 0.001)$, indicating a constant decrease in the dependent variable with increasing price, the quadratic model fails to provide additional explanatory power, with $non-significant$ coefficients for both the linear and squared terms. The significance of the cubic term in the most complex model suggests a$non-linear$ relationship, potentially involving inflection points where the impact of price changes varies across its range. This result aligns with prior research on $non-linear$ pricing effects, emphasizing the need for higher-order terms to capture complex economic dynamics. Further research should explore these $non-linear$ patterns with larger datasets and additional variables to validate and generalize these findings.

The regression analysis illustrates a negative linear relationship between the price of tomato paste and the dependent variable, as shown by the unstandardized coefficient (B = -0.022). This indicates that for every unit increase in the price of tomato paste, the dependent variable is expected to decrease by 0.022 units. However, this relationship is not statistically significant $(p = 0.276)$, suggesting that the price of tomato paste does not reliably predict changes in the response variable within this linear model. The constant term $(B = 169.204)$, representing the predicted value of the response variable when the price of tomato paste is zero, is statistically significant $(t = 4.836, p = 0.000)$. Despite the apparent trend, the lack of significance in the linear relationship undermines its predictive validity, indicating that alternative models or variables may better explain the observed variation.

3.2 Discussion of Results of Sale Oil Models:

Table 3 illustrates the ANOVA results of oil sales. It can be noticed that the Quadratic Model has the best fit since its mean square error is smaller, MSE: 4598.188, and the standard error: 69.038, compared to the Linear Model with a MSE of 4766.225 and Std. Error of 69.038, and the Cubic Model, which is 4567.330 and 67.582, respectively. Fig. 3 compares the performance of linear, quadratic, and cubic regression models in capturing the relationship between the price of oil and the dependent variable. The quadratic and cubic models both exhibit statistically significant fits $(p = 0.021)$, indicating that they can capture underlying non-linear patterns in the data. In contrast, the Linear model, with a p-value of 0.276, fails to achieve statistical significance, suggesting it is insufficient in describing the observed relationship. Among the models, the quadratic model strikes a balance between simplicity and accuracy, effectively capturing the trends with minimal overfitting. While the cubic model adds complexity, it does not significantly outperform the quadratic model, emphasizing that the quadratic model is likely the most appropriate choice for generalization and interpretability in this

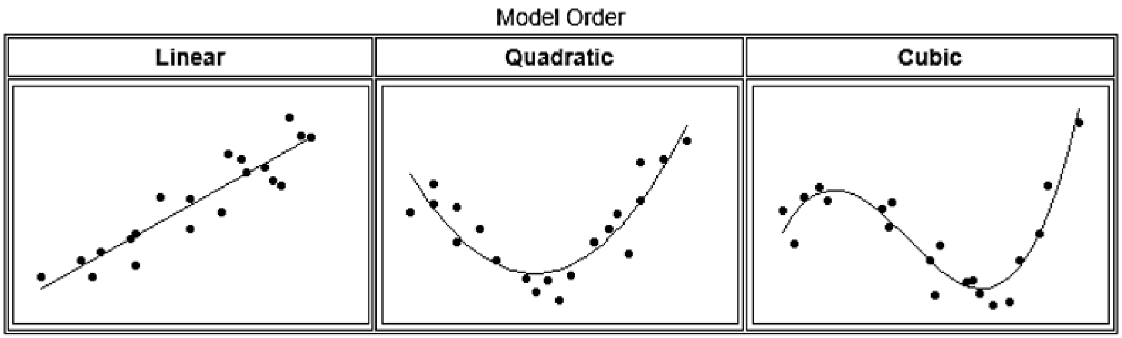

Figure 1. Different orders models of regression analysis

Table 1. The results of different models of the ANOVA table of tomato paste sales

Model	Mean Squared Error	F	Sig.	Std. Error of the Estimate
Linear Regression	3635.405	11.354	.001	60.294
Quadratic	3643.968	5.982	.003	60.365
Cubic	3643.968	5.982	.003	60.365

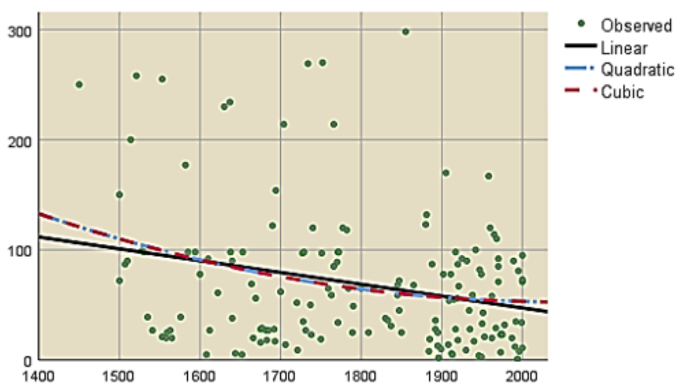

Figure 2. Models of selling the tomato paste.

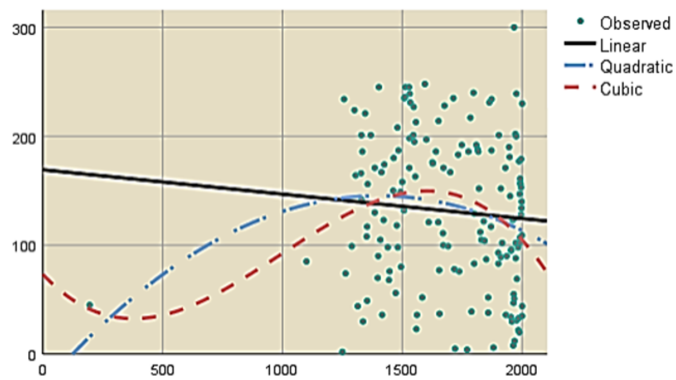

Figure 3. Models of the oil sales.

analysis. The analysis of the relationship between the price of X and the response variable reveals distinct patterns across the three models. The linear regression model indicates a negative but statistically insignificant relationship (p=0.276), suggesting it does not adequately capture the trend. In contrast, the quadratic model identifies a significant positive relationship (p=0.021), along with a diminishing effect as the price of X increases, evidenced by the significant quadratic term (p=0.011). This suggests that the quadratic model provides a better fit, capturing the non-linear dynamics of the data. However, the cubic model, despite its added complexity, fails to yield significant coefficients ($p > 0.05$) and does not improve interpretability over the quadratic model. Thus, the quadratic model emerges as the most suitable for explaining the relationship between the Price of X and the dependent variable.

The following three fitted models linear, quadratic, and cubic are shown in different colors along with the price of oil against the variation of the dependent variable throughout the observed data points for the same. The least square linear model is in black and has a slight negative slope, which indicates a relationship that is inverse and weak and does not capture the obvious curvature or trend in the data.

This shows a very good fit for the quadratic model, as is observed by the blue curve in the figure below. It shows that with an increase in the price of oil, the dependent variable goes up and reaches a maximum at middle-range prices, roughly in the range of 1000-1500 then turns to decline with increased prices-a diminishing the effect of higher-range price increases.

Table 2. The different results of the estimated tomato paste sales

Model	Variable	Unstandardized Coeff.		t	Sig.
		B	Std.Error		
Linear Regression	Price of X	-.107	.032	-3.370	.001
	(Constant)	261.721	57.346	4.564	.000
Quadratic	Price of X	-.783	.848	-.923	.357
	Price of X^2	.000	.000	.797	.426
	(Constant)	854.174	745.261	1.146	.254
Cubic	Price of X	-.783	.848	-.923	.357
	Price of X^2	.000	.000	.797	.426
	Price of X^3	-81.906	.858	-2.070	.040
	(Constant)	854.174	745.261	1.146	.254

Table 3. The different results of the ANOVA table of oil sales models

Model	Mean Squared Error	F	Sig.	Std. Error of the Estimate
Linear Regression	4766.225	1.197	0.276	69.038
Quadratic	4598.188	3.252	0.021	67.810
Cubic	4567.330	3.333	0.021	67.582

The cubic model, represented by the red curve, adds even more undulations and improves the fit only a little. This added complication doesn't help much with interpretability but makes the model prone to fitting the noise in the data.

Overall, the quadratic is the best fit. It captured the general trend: the positive influence of price on the dependent variable up to a point - somewhere between 1000 and 1500 - after which it decreases. This suggests that moderate prices are optimal for higher values of the dependent variable, whereas very low or very high prices are not as good.

4. Conclusions:

The linear regression model of Tomato Paste provides the most effective fit for predicting the price of tomato paste. It demonstrates a slightly better overall fit based on the F-statistic and p-values, which indicate a stronger model performance. The model also shows nearly identical Mean Squared Error (MSE) and Standard Error of Estimate (SEE), suggesting that the linear model delivers comparable prediction accuracy and error margins when compared to other models. For the price of X, the linear regression model is the most effective due to the statistical significance of the variable Price of X. This model presents a clear, interpretable relationship between the independent and dependent variables. While the quadratic and cubic models include non-linear terms, only the cubic model showed statistical significance for Price of X^3. However, the linear model still stands out as the best choice for simplicity and model fit.

When predicting the price of oil, the cubic regression model outperforms both the linear and quadratic models. It achieves the best prediction accuracy with the lowest SEE, making it the most effective model for capturing the complex relationship between oil price and sales. The quadratic model is the most effective for explaining the price of oil. Both Price of X and its squared term (Price of X^2) are statistically significant, indicating a non-linear, parabolic relationship between oil price and the dependent variable. This suggests that oil prices influence sales in a non-linear fashion, with the price's impact changing at different levels.

5. Recommendations:

1. The quadratic and cubic models, even though they are statistically significant for the price of tomato paste, do not offer any extra predictive power or improvement in error metrics when compared to the linear model. Therefore, you might consider using the linear model for its simplicity and effectiveness unless there is a specific reason to use more complex models, such as theoretical expectations of non-linearity in the data.

2. The cubic model is the most successful in explaining the variation and delivering accurate forecasts, given that the MSE and SEE for the price of oil are lower than

Table 4. The different results of the estimated oil sales models

Model	Variable	Unstandardized Coeff.		t	Sig.
		B	Std.Error		
Linear Regression	Price of X	-.022	.020	-1.094	0.276
	(Constant)	169.204	34.989	4.836	0.000
Quadratic	Price of X	.250	.107	2.327	0.021
	Price of X^2	-8.908E-5	.000	-2.582	0.011
	(Constant)	-29.504	84.296	-.350	0.727
Cubic	Price of X	-.233	.354	-.657	0.512
	Price of X^2	.000	.000	1.151	0.252
	Price of X^3	-1.278E-7	.000	-1.428	0.155
	(Constant)	73.190	110.574	.662	0.509

the other models. If you believe the data might have a more complex non-linear relationship, the cubic model would be the optimal choice.

Funding: There is no financial support or sponsorship from any institute.

Data Availability Statement: All of the data supporting the findings of the presented study are available from corresponding author on request.

Declarations:

Conflict of interest: The authors declare that they have no conflicts of interest.

Ethical approval: The Declaration of Helsinki's ethical guidelines guided the research. We conducted the procedure after obtaining the premonition from the market known as Golden Company in Erbil, Iraq for the production and marketing of tomato paste & oil that has been experienced in this field for many years, then the data investigates the price sales of two items and their effects on the number of items sales in the market.

Author contributions: Hemn Hussein Yaseen collected the samples or data, conducted the analysis, interpreted the data, drafted the manuscript, and performed proofreading. Paree khan Abdulla Omer conceived the research idea, supervised the paper, and reviewed the manuscript."

References

[1] Julie E Adams, Randall W Culp, and A Lee Osterman. Interosseous membrane reconstruction for the essex-lopresti injury. *The Journal of hand surgery*, 35(1):129–136, 2010, doi:10.1016/j.jhsa.2009.10.007.

[2] Pete Smith, Christian A Davies, Stephen Ogle, Giuliana Zanchi, Jessica Bellarby, Neil Bird, Robert M Boddey, Niall P McNamara, David Powlson, Annette Cowie, et al. Towards an integrated global framework to assess the impacts of land use and management change on soil carbon: current capability and future vision. *Global Change Biology*, 18(7):2089–2101, 2012, doi:10.1111/j.1365-2486.2012.02689.x.

[3] Amy E Taylor, Hannah J Jones, Hannah Sallis, Jack Euesden, Evie Stergiakouli, Neil M Davies, Stanley Zammit, Debbie A Lawlor, Marcus R Munafò, George Davey Smith, et al. Exploring the association of genetic factors with participation in the avon longitudinal study of parents and children. *International journal of epidemiology*, 47(4):1207–1216, 2018, doi:10.1093/ije/dyy060.

[4] MO Clarkson, RA Wood, Simon W Poulton, S Richoz, RJ Newton, Simone A Kasemann, F Bowyer, and L Krystyn. Dynamic anoxic ferruginous conditions during the end-permian mass extinction and recovery. *Nature Communications*, 7(1):12236, 2016.

[5] Douglas H Smith, Patrick M Kochanek, Susanna Rosi, Retsina Meyer, Chantelle Ferland-Beckham, Eric M Prager, Stephen T Ahlers, and Fiona Crawford. Roadmap for advancing pre-clinical science in traumatic brain injury. *Journal of Neurotrauma*, 38(23):3204–3221, 2021, doi:10.1089/neu.2021.0094.

[6] Donald M Lloyd-Jones, Pamela B Morris, Christie M Ballantyne, Kim K Birtcher, David D Daly, Sondra M DePalma, Margo B Minissian, Carl E Orringer, and Sidney C Smith. 2017 focused update of the 2016 acc expert consensus decision pathway on the role of non-statin therapies for ldl-cholesterol lowering in the management

of atherosclerotic cardiovascular disease risk: a report of the american college of cardiology task force on expert consensus decision pathways. *Journal of the American College of Cardiology*, 70(14):1785–1822, 2017, doi:10.1016/j.jacc.2016.03.519.

[7] Eva Ostertagová. Modelling using polynomial regression. *Procedia engineering*, 48:500–506, 2012, doi:10.1016/j.proeng.2012.09.545.

[8] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning: with applications in R*, volume 103. Springer, 2013, doi:10.1007/9.

[9] Dhruv Grewal, Scott Motyka, and Michael Levy. The evolution and future of retailing and retailing education. *Journal of Marketing Education*, 40(1):85–93, 2018, doi:10.1177/0273475318755838.

[10] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

[11] Rajendra Kunwar and GC Laxmi. An overview of strategies for conceptualizing derivative and their applications in daily life for secondary-level mathematics students. *South Asian Res. J. Eng. Tech.*, 5:69–82, 2023, doi:10.36346/sarjet.2023.v05i05.003.

[12] James Stewart, Saleem Watson, and Daniel Clegg. Calculus: Early transcendentals, 8th. *Edition, Brooks/Cole, Cengae learning*, 4:13, 2012.

[13] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021, doi:10.1561/1100000095.

[14] Syed Saad Andaleeb and Khalid Hasan. *Strategic marketing management in Asia: case studies and lessons across industries*. Emerald Group Publishing Limited, 2016, doi:10.1108/978-1-78635-746-520161020.

[15] Joseph SW Lewis, Pierre Ocvirk, Dominique Aubert, Jenny G Sorce, Paul R Shapiro, Nicolas Deparis, Taha Dawoodbhoy, Romain Teyssier, Gustavo Yepes, Stefan Gottlöber, et al. Galactic ionizing photon budget during the epoch of reionization in the cosmic dawn ii simulation. *Monthly Notices of the Royal Astronomical Society*, 496(4):4342–4357, 2020, doi:10.1093/mnras/staa1748.

[16] L Novotnỳ. Utilization of finite element method in deformation calculation of assembly platform. *Acta Mechanica Slovaca*, pages 607–612, 2008, doi:10.1016/j.proeng.2012.09.545.

[17] Trebuňa F. Frankovský P. amp; Binda M. Masláková, K. Applications of the strain gauge for determination of residual stresses using ring-core method. *method. Procedia Engineering, 48,*, 48:396–401, 2012,doi:10.1016/j.proeng.2012.09.531.

[18] P Frankovskỳ, A Kostelníková, and P Šarga. The use of strain-gage method and photostress method in determining residual stresses of steel console. *Metalurgija*, 49(2):208–212, 2010, doi:10.4028/www.scientific.net/AMM.732.3.

التحليل التنبؤي لمبيعات السوق باستخدام نماذج الانحدار متعدد الحدود

همن حسن ياسين [1*] ، بری خان عبالله عمر [2]

[1] قسم الإحصاء والمعلوماتية، كلية الإدارة والاقتصاد، جامعة صلاح الدين، اربيل، العراق.

* الباحث المسؤول: *hemnblbas95@gmail.com*

الخلاصة

في هذه الورقة، نقوم بتحليل بيانات المبيعات لعنصرين مختارين من سوبر ماركت ونطبق نماذج الانحدار متعدد الحدود للكشف عن رؤى حول اتجاهات مبيعاتهما. من خلال ملاءمة منحنى متعدد الحدود للبيانات، نهدف إلى فهم كيفية تأثير هذا التسعير على عدد مبيعات هذه العناصر. يقارن هذا التحليل فعالية نماذج الانحدار المختلفة في التنبؤ بسعر المنتجات في سياقات مختلفة. نموذج الانحدار الخطي هو الأكثر فعالية للتنبؤ بسعر معجون الطماطم، لأنه يمنح البيانات ملاءمة أفضل قليلاً، مع *MSE* و *SEE* متطابقين تقريبًا، مما يشير إلى دقة وأخطاء تنبؤ مماثلة. نموذج الانحدار الخطي هو الأكثر فعالية للتنبؤ بالمتغير التابع وهو عدد مبيعات عنصر معجون الطماطم، مما يسلط الضوء على أن سعر العنصر (المتغير المستقل) ذو دلالة إحصائية ويوفر علاقة واضحة مع المتغير التابع. في سياق التنبؤ بعدد سلع الزيت كمتغير تابع، فإن النموذج التربيعي هو الأكثر فعالية في تفسير سعر سلعة الزيت مع كون المتغيرين المستقلين (سعر القوة واحد وسعر القوة اثنين) مهمين إحصائيًا، مما يشير إلى أن النموذج التربيعي يصف بشكل أفضل العلاقة بين المتغيرين.

الكلمات الدالة : نموذج الانحدار الخطي المتعدد، نموذج الانحدار متعدد الحدود، اقل المربعات خطأ

التمويل: لايوجد.

بيان توفر البيانات: جميع البيانات الداعمة لنتائج الدراسة المقدمة يمكن طلبها من المؤلف المسؤول.

اقرارات:

تضارب المصالح: يقر المؤلفون أنه ليس لديهم تضارب في المصالح.

الموافقة الأخلاقية: لم يتم نشر المخطوطة أو تقديمها لمجلة أخرى، كما أنها ليست قيد المراجعة.

مساهمات المؤلفين: قام هيمن حسين ياسين بجمع العينات أو البيانات، وإجراء التحليل، وتفسير البيانات، وصياغة المخطوطة، ومراجعتها. أما بري خان عبد الله عمر، فقد ابتكرت فكرة البحث، وأشرفت على الورقة، وراجعت المخطوطة.