International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 10 October. 2026, pp:242-248 www.ijaem.net ISSN: 2395-5252

Conditional Distribution of Oat Harvested Area and Production Quantity: A Comparison of OLS and Quantile Regression

Şenol Çelik[1]
[1]Unit of Biometrics and Genetics, Department of Animal Science, Faculty of Agriculture, Bingöl University, Bingöl-Türkiye

Date of Submission:09-10-2026 Date of Acceptance: 22-10-2026

Abstract: In this study, the relationship between oat production quantity and harvested area across the districts of Konya in 2025 is analyzed using both classical Ordinary Least Squares (OLS) regression and the Quantile Regression model, which captures structural variations across different points of the conditional distribution. Utilizing data from 27 districts, bootstrap standard errors were applied to ensure parameter stability, and model fit was evaluated using Koenker-Machado Pseudo-R^2 measures. While classical OLS results indicate a positive and significant average effect of harvested area on production ($\beta = 0.248$), quantile regression reveals that this marginal effect varies considerably across production scales. The findings demonstrate that the marginal effect is lower in the lower quantile ($\tau = 0.25, \beta = 0.090$) and reaches its peak in the upper quantile ($\tau = 0.75, \beta = 0.329$). These results confirm that heterogeneity arising from scale economies and regional productivity differences in agricultural data is masked by OLS, whereas quantile regression provides a superior and comprehensive approach for examining agricultural production dynamics.

Keywords: Quantile Regression, Oat Production, Konya, Scale Heterogeneity.

I. Introduction

The most extensive use of oats (*Avena sativa* L.) is in animal nutrition. It is also a cereal crop used for human consumption (Buerstmayr, 2007). Due to its high protein content, oats are an important crop for feeding livestock (Wood, 2001). While oats intended for human consumption are valued for their high levels of protein, β-glucan, and soluble fiber along with a low fat content oats used for animal feed are ideally required to have high protein and fat content, and a low β-glucan content, in order to provide maximum energy (Peterson et al., 2005).

In both Türkiye and the rest of the world, the growing awareness of healthy nutritiondriven by rising income and prosperity levelshas brought the consumption of oats to the forefront. Research indicates that oats contain antioxidants, including simple phenolic compounds (ferulic, caffeic, p-coumaric, sinapic, and vanillic acids) (White and Xing, 1997) as well as avenanthramides (Dokuyucu et al., 2003). Furthermore, their high fiber contentwhich is effective in lowering cholesteroland their iron content make them an excellent choice for a healthy diet (Wood, 2001).

According to 2025 data from the Turkish Statistical Institute (TSI), the province of Konya ranks third in Türkiye for oat production. The provinces with the highest oat production are, in order, Sivas, Ankara, and Konya. The oat production amounts for these provincesmeasured in tonsare 42,017, 28,397, and 21,261, respectively.

There are various studies in the literature concerning production forecasting for food products. Wenjun et al. (2007) and Olowa and Olowa (2016) employed the linear regression method to forecast demand for agricultural, fishery, and livestock products. Wang et al. (2024) used artificial neural networks to estimate the production volumes of fish and rice in China, while Kharal et al. (2024) attempted to predict wheat yield using support vector machines.

Traditional econometric studies in the literature generally employ the Ordinary Least Squares (OLS) regression method to examine the "average" relationship between variables (Gujarati, 2021). While the OLS model is effective at estimating the conditional mean of a dependent variable, it proves inadequate in the presence of heteroscedasticity (varying variance) or outliers, and it tends to obscure specific dynamics occurring at the lower or upper ends of the distribution (Koenker& Bassett, 1978). In the context of agricultural production data, however, the marginal effects and the corresponding responses differ significantly between districts with low production levels and those with massive production capacities. In this regard, the Quantile Regression method originally developed by Koenker and Bassett (1978)offers a more comprehensive and robust approach by focusing not merely on the mean of the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 10 October. 2026, pp:242-248 www.ijaem.net ISSN: 2395-5252

data distribution, but on various quantiles (τ), including the median (Koenker, 2007).

The primary objective of this study is to examine the relationship between oat production amount and harvested area across the districts of Konya province for the year 2025, utilizing a quantile regression approacha method that extends beyond the conventional ordinary least squares (OLS) model. The study empirically tests whether the marginal contribution of the harvested area to production varies according to the production scale of each district, while also identifying patterns of regional heterogeneity. In this regard, the research aims to provide a methodological contribution to the agricultural economics literature and to offer guidance for agricultural planning policies at the local level.

II. Material and Methods

Material

The primary material for this study consists of official, district-level data on agricultural production activities in the districts of Konya province for the year 2025. The analysis covers the oat production amount (in tons) and the corresponding harvested area (in decares/hectares) for the 27 districts within Konya province. The data were compiled from agricultural statistics and district-based agricultural inventory records (TSI, 2025). In this research, the amount of oat production is used as the dependent variable (Y), while the harvested area serves as the independent variable (X).

Method

In this study, both Ordinary Least Squares (OLS) regression and Quantile Regression models were applied to examine the impact of the harvested area at the district level on oat production, and to reveal potential structural heterogeneity within the data.

Linear Regression (OLS) Model

The general equation for an OLS model, constructed to estimate the conditional mean of a dependent variable, is as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

In this model, $Y_i$represents the oat production quantity in ith district; $X_i$represents the harvested area; β_0 is the constant term; β_1 is the average marginal effect of the harvested area on production; and $\varepsilon_i$is the error term with a mean of zero (Gujarati, 2021).

Quantile Regression Model

Since classical OLS regression focuses solely on the mean relationship, it may obscure asymmetries such as those arising from heteroscedasticity (non-constant variance) and scale differences that are frequently encountered in agricultural production data. To overcome this limitation, the quantile regression method, developed by Koenker and Bassett (1978), has been adopted. Quantile regression allows for the modeling of any given τ th quantile $(0 < \tau < 1)$ of the conditional distribution of a dependent variable as a linear function of the independent variable:

$$Q_Y(\tau \mid X_i) = \beta_0(\tau) + \beta_1(\tau)X_i$$

Here, $Q_Y(\tau \mid X_i)$represents the value of the τ-th quantile of Y given X_i, while $\beta_0(\tau)$ and $\beta_1(\tau)$ denote the parameter estimates at the corresponding quantile levels $(\tau = 0.25, 0.50, 0.75)$ (Koenker, 2007). The expression for quantile regression in terms of error terms (a framework that minimizes the weighted sum of absolute errors) is as follows:

$$Y_i = \beta_0(\tau) + \beta_1(\tau)X_i + \varepsilon_i(\tau)$$

In this study, the relationship between the quantity of oat production and the harvested area at the district level was estimated using the standard Ordinary Least Squares (OLS) model,$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$;furthermore, to observe structural variations at different points of the distribution, quantile regression modelsexpressed by the formulation $Q_Y(\tau \mid X_i) = \beta_0(\tau) + \beta_1(\tau)X_i$were developed (for $\tau = 0.25, 0.50, 0.75$).

Given the small sample size (n=27), standard errors were calculated using the bootstrap method to ensure the statistical stability and reliability of the parameter estimates.

Model Fit Criteria and Tests

To evaluate the explanatory power of the models, the coefficient of determination (R^2) was used for the OLS model, while the Pseudo-R^2 measuredeveloped by Koenker and Machado (1999)was employed for the quantile regression models. Furthermore, the Koenker-Bassett equality test was applied to determine whether the slope coefficients across different quantiles differ from one another at a statistically significant level. All analyses were conducted using the R (v.4.x) software and the `quantreg` package (R Core Team, 2025).

III. Results

In this study, the relationship between oat production volume and harvested area in the districts of Konya province for the year 2025 was analyzed using Ordinary Least Squares (OLS) regression and Quantile Regression models, which examine structural variation across different segments of the distribution.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 10 October. 2026, pp:242-248 www.ijaem.net ISSN: 2395-5252

OLS Regression Model Findings

The estimation results from the classical OLS model, based on district-level data, indicate that the harvested area has a positive and statistically highly significant effect on oat production ($p<0.001$). The coefficient of determination (R^2) for the model is calculated as 0.813 (Adjusted R^2 = 0.8056). This value reveals that approximately 81.3% of the total variance in the quantity of oat production across the Konya province is explained by the harvested area. According to the OLS model, each 1-unit increase in the harvested area leads, on average, to a 0.248-unit increase in the quantity of oat production ($t=10.427$).The mathematical expression of the OLS model, using the obtained coefficients, is:

$$Y_i = -66.86 + 0.248X_i + \varepsilon_i \quad (R^2 = 0.813)$$

Findings from the Quantile Regression Model

The quantile regression results indicate that the association between harvested area and oat production varies across the conditional distribution of production in the 27 districts of Konya Province. The estimated coefficient of harvested area increases from 0.090 at the 25th quantile to 0.247 at the median and 0.329 at the 75th quantile. These estimates suggest that the relationship between harvested area and production may be stronger among districts located in the upper part of the conditional production distribution.

The analysis results demonstrate that the marginal effect of the harvested area on production varies significantly depending on the production scale of the districts:

Lower Quantile (τ=0.25): At the 25th quantile, the estimated coefficient is relatively small (β = 0.090; bootstrap SE = 0.074), indicating substantial uncertainty regarding the association at this quantile. At the median, the coefficient (β = 0.247; bootstrap SE = 0.070) is close to the OLS estimate of 0.248. At the 75th quantile, the coefficient is higher (β = 0.329; bootstrap SE = 0.047), suggesting a stronger positive association between harvested area and oat production at higher conditional production levels.

$$Q_{\text{production}}(0.25 \mid \text{Harvest}_i) = 2.98 + 0.090 \cdot \text{Harvest}_i \quad (\text{Pseudo-}R^2 = 0.387)$$

$$Y_i = 2.98 + 0.090\ X_i$$

In districts located in the lower segment of the production distribution, the marginal contribution of the harvested area to production is 0.090.

Median Quantile (τ=0.50): In the median model, which represents the central tendency, the coefficient for the harvest area was found to be 0.247 ($p=0.0016$); this result shows a high degree of alignment with the mean coefficient obtained from the OLS regression. The Pseudo-R^2 value for this model is 0.614.

$$Q_{production}(0.50 \mid \text{Harvest}_i) = -5.36 + 0.247 \cdot \text{Harvest}_i \quad (\text{Pseudo-}R^2 = 0.614)$$

$$Y_i = -5.36 + 0.247X_i$$

The relationship in the median district shows a strong fit with the OLS model, revealing that every 1-unit increase in area contributes 0.247 units to production.

Upper Quantile (τ=0.75): In large-scale districts with high production potential, the marginal effect of the harvested area on production reaches its peak, with a value of 0.329 ($p<0.001$). In this upper segment, the explanatory power of the modelindicated by a Pseudo-R^2 of 0.805comes very close to the R^2value of the OLS model.

$$Q_{\text{production}}(0.75 \mid \text{Harvest}_i) = -6.31 + 0.329\text{Harvest}_i \quad (\text{Pseudo-}R^2 = 0.805)$$

$$Y_i = -6.31 + 0.329X_i$$

In the top tier of districtsthose with high production potentialthe marginal effect of the harvested area on production reaches its peak at 0.329. This mathematically demonstrates that the conversion of land area into production operates much more efficiently (or robustly) in districts where large-scale agriculture is practiced.

Model Comparison and Summary Table

The OLS and quantile regression parameter estimates, standard errors, and goodness-of-fit criteria obtained in the study are presented collectively in Table 1.

Table 1. OLS and Quantile Regression Estimation Results for Oat Production in Konya Province

Variable	OLS (Mean)	Quantil($\tau = 0.25$)	Median ($\tau = 0.50$)	Quantil($\tau = 0.75$)
Intercept	-66.86 (156.36)	2.98 (26.33)	-5.36 (29.14)	-6.31 (19.84)
Harvest Area (X)	0.248*** (0.024)	0.090 (0.074)	0.247*** (0.070)	0.329*** (0.047)
	$R^2 = 0.813$	Pseudo-R^2 = 0.387	Pseudo-R^2 = 0.614	Pseudo-R^2 = 0.805

(Note: The values in parentheses below the parameter estimates represent Bootstrap standard errors; *** indicates statistical significance at the $p<0.001$ level.)

International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 10 October. 2026, pp:242-248 www.ijaem.net ISSN: 2395-5252

When the findings in Table 1 are evaluated as a whole, it is observed that the ratio of harvest area to production increases exponentially as one moves from the lower quantiles to the upper quantiles. This empirically confirms the existence of a structural heterogeneity across the districts of Konya, stemming from agricultural economies of scale and regional differences in productivity.

When the parameter estimates presented in Table 1 are examined as a whole, it becomes clear that while the mean value provided by the OLS model (β=0.248) successfully reflects the general trend of the data (R^2=0.813), it masks structural differences at the lower and upper ends of the distribution. Quantile regression models reveal this hidden heterogeneity across the following three key dimensions:

1. Economies of Scale and Productivity Disparity: The harvested area coefficient, which stands at 0.090 in the lower quantile (τ=0.25) model, indicates that the yield obtained per unit of land remains limited in districts with low production amounts or small-scale agricultural operations. In contrast, the increase of this coefficient to 0.329 in the upper quantile (τ=0.75) model suggests that the conversion of land into production is significantly more efficient and robust in districts characterized by large-scale agricultural operations (such as Yunak, Tuzlukçu, and Ilgın).

2. Variation in Model Fit: The Koenker-Machado Pseudo-R^2 values increase as one moves from the lower quantile to the upper quantile (from 0.387 to 0.805). This demonstrates that the model's explanatory power and predictive capability are significantly higher for districts with high production amounts.

3. Methodological Superiority: The fact that the coefficient in the median model (τ=0.50)which is 0.247remains quite close to the OLS estimate confirms the median's effectiveness in capturing central tendency; meanwhile, the deviations observed at the lower and upper quantiles provide empirical validation for the appropriateness and necessity of employing the quantile regression method in this study.

Due to scale differences and heteroskedasticity observed in district-level agricultural production data, the Quantile Regression modelwhich goes beyond the OLS model by providing a comprehensive view of marginal effects across different segments of the distributionhas been determined as the most appropriate and inclusive analytical approach for this study.

The reported Koenker–Machado pseudo-R^2 values are 0.387, 0.614, and 0.805 for the 25th, 50th, and 75th quantiles, respectively. These values indicate differences in model fit across the examined quantiles. However, pseudo-R^2 values from quantile regression are not directly comparable to the conventional OLS R^2, and they should not be interpreted as direct evidence of superior predictive performance.

Overall, the findings suggest heterogeneity in the relationship between harvested area and oat production across the conditional production distribution. Nevertheless, the observed differences between quantile-specific coefficients should be formally tested before concluding that the association differs significantly across quantiles. Given the relatively small sample size (n = 27), the uncertainty of the bootstrap estimates should also be considered when interpreting the results. Furthermore, the cross-sectional nature of the data does not establish a causal relationship or, by itself, demonstrate economies of scale or differences in technical efficiency.

Quantile regression coefficient estimates and bootstrap confidence intervals for oat production in the province of Konya are shown in Figure 1.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 10 October. 2026, pp:242-248 www.ijaem.net ISSN: 2395-5252

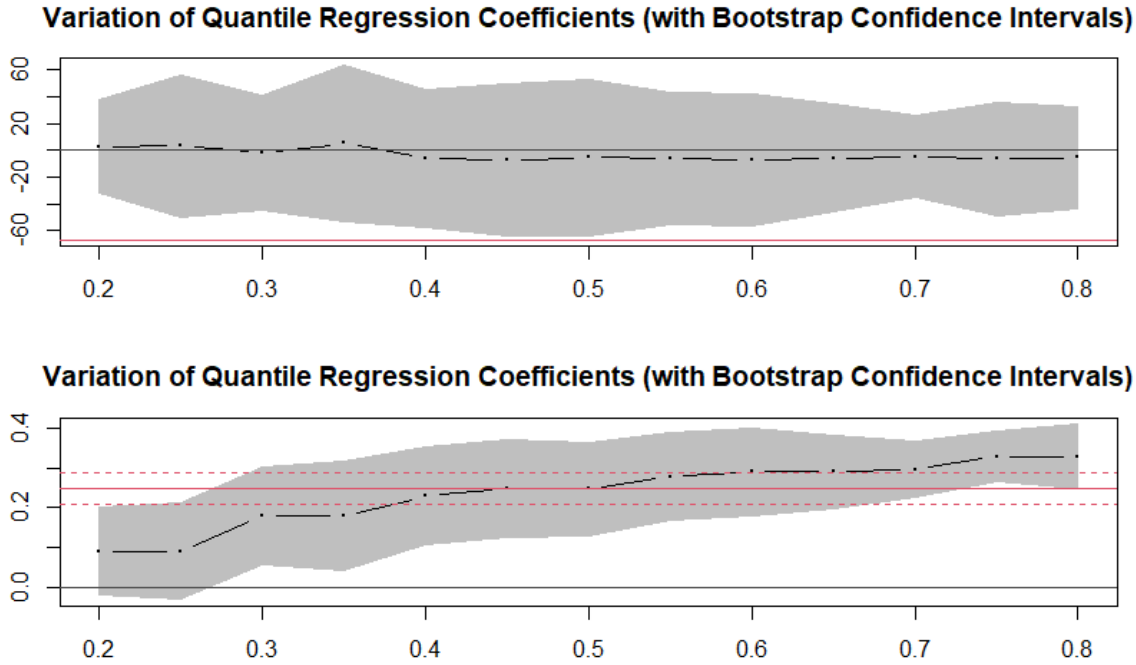

Figure 1. Quantile Regression Coefficient Estimates and Bootstrap Confidence Intervals for Oat Production in Konya Province

As shown in Figure 1, the black dashed line in the bottom panel traces the coefficients for the harvested area across various quantile levels (ranging from τ=0.20 to 0.80). The solid red horizontal line and the dashed red lines represent the mean coefficient and the confidence interval, respectively, as estimated by the Ordinary Least Squares (OLS) model. At lower quantiles (τ=0.20 and 0.25), the coefficients are around 0.09; however, as the quantile level increases, the coefficients show a consistent upward trend, reaching approximately 0.33 at the higher quantiles (τ=0.75 and 0.80). This graph demonstrates that not all districts in Konya share the same production dynamics for oat cultivation. In districts situated in the lower segmentcharacterized by lower yields or smaller scalesthe marginal contribution of the harvested area to total production is relatively limited, whereas in districts in the upper segmentthose with higher potentialthe yield per unit area is significantly greater. Although the red OLS line in the graph provides a reasonable representation of the center of the distribution (near the median), it obscures this structural variation (asymmetry) present in the lower and upper segments. This underscores both the validity and the necessity of employing quantile regression for this dataset.

IV. Discussion

In this study, the relationship between oat production volume and harvested area across the districts of Konya province for the year 2025 was examined using both an OLS model and a quantile regression modelthe latter being capable of revealing structural heterogeneity across different segments of the distribution. The empirical findings demonstrate that the relationship between agricultural production and land area is neither linear nor homogeneous; rather, it exhibits significant variations depending on the production scale of the individual districts.

In one study, the Moving Average model also yielded the lowest MSE values for oat and corn production forecasts, demonstrating that it provides the most accurate predictions for the production of these crops (Nalici et al., 2024).

In one of the studies conducted on the oat plant, it was observed that in the first year, the genotypes exhibited higher values for grain yield, plant height, protein content, and NDF, whereas in the second year, 1000-grain weight and starch content were higher. There were no significant differences among the genotypes in terms of hectoliter weight and ADF (Kahraman and Alatürk, 2024).

From a methodological perspective, this study reaffirms the utility of the quantile regression approachas highlighted by Koenker and Bassett

(1978)due to its robustness against outliers and its capacity to provide detailed insights into the entire distribution. The findings suggest that agricultural planning and support policies should not be applied uniformly (i.e., based on averages) across all districts; rather, it is essential to support small-scale producers with inputs that enhance productivity, while simultaneously preserving the sustainable economies of scale in large-scale production centers.

V. Conclusion

In this study, the relationship between oat production volume and harvested area in the province of Konya was examined using 2025 district-level data, employing both OLS regression and a quantile regression model that comprehensively addresses distributional dynamics. The key findings of the research can be summarized as follows:

1. Determination of the Average Trend: The OLS model confirmed that the harvested area has a generally positive, strong, and statistically significant effect on oat production (R^2=0.813, β=0.248).

2. Evidence of Scale Heterogeneity: Quantile regression analyses reveal that the relationship between agricultural land area and production is not homogeneous; rather, districts exhibit distinct structural differences based on their production scale. The marginal effect of harvested area on production is 0.090 at the lower quantiles (τ=0.25), rising to 0.329 at the upper quantiles (τ=0.75). This demonstrates that the conversion of land area into production operates with significantly greater efficiency in districts characterized by large-scale agricultural operations.

3. Policy Recommendations: The findings suggest that agricultural support and planning policies should not rely on uniform (average-based) strategies; instead, they should be designed with due consideration for regional variations in scale and productivity. In districts with small-scale production, it is essential to increase incentives for productivity-enhancing inputs and modern agricultural techniques, while in large-scale production hubs, the priority should be to preserve existing economies of scale and ensure long-term sustainability.

It is considered that in future studies, regional heterogeneity in agricultural production could be addressed through more comprehensive models by incorporating additional inputssuch as fertilizers, pesticides, irrigation facilities, and climatic datainto the analysis.

References

[1]. Buerstmayr H, Krenn N, Stephan U, Grausgruber H, Zechner E. 2007. Agronomic Performance and Quality of Oat (Avena sativa L.) Genotypes of Worldwide Origin Produced Under Central European Growing Conditions. Field Crops Res. 101: 341-351.

[2]. Dokuyucu T, Peterson DM, Akkaya A. 2003. Contents of Antioxidant Compounds in Turkish

[3]. Oats: Simple Phenolics and Avenanthramide Concentrations, Cereal Chem., 80(5): 542-543.

[4]. Gujarati, D. N. 2021. Temel Ekonometri (Çev. Ed. Ü. Şenesen& M. Şenesen). Literatür Yayıncılık.

[5]. Kahraman, T., Alatürk, F. 2024. Farklı Yulaf (Avenasativa L.) Genotiplerinin Tane Verimi ve Yem Kalitesi Yönünden Değerlendirilmesi. Bilim-Teknoloji-Yenilik Ekosistemi Dergisi, 5(2), 81-89.

[6]. Kharal, A. S., Mahar, S. A., Mushtaque, M. I., Magsi, A., Mahar, J. A. 2024. A model for wheat yield prediction to reduce the effect of climate change using support vector regression. VFAST Transactions on Software Engineering, 12(2), 192-212.

[7]. Koenker, R. 2007. QuantileRegression. Cambridge UniversityPress, xvi + 350 ISBN 0-521-60827-9

[8]. Koenker, R., Bassett, G. 1978. Regressionquantiles. Econometrica, 46(1), 33–50. https://doi.org/10.2307/1913643

[9]. Koenker, R., Machado, J. A. F. 1999. Goodness of fit andrelatedinferenceforquantile regression. Journal of theAmerican Statistical Association, 94(448), 1296–1310. https://doi.org/10.1080/01621459.1999.10473882

[10]. Nalici, M. E., Soylemez, İ., Ünlü, R. 2024. Türkiye'deTahılÜretimininTahminlemesi: Karşılaştırmalı Analiz. Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi, 40(2), 288-303.

[11]. Olowa, O. W., Olowa, O. A. 2016. Food demand forecast for Nigeria (2016–2028).Scientia Agriculturae, 15(1), 340–343.

[12]. Peterson DM, Wesenberg, DM, Burrup DE, Erickson CA. 2005. Relationships among Agronomic Traits and Grain Composition in Oat Genotypes Grown in Different Environments. Crop Sci., 45, 1249-1255

[13]. R Core Team. 2025. R: A languageandenvironmentforstatisticalcomputing. RFoundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
[14]. TÜİK, (TSI) (2025). Tahıllar ve diğer bitkisel ürünler. Türkiye İstatistik Kurumu.
[15]. https://biruni.tuik.gov.tr/medas/?kn=92&locale=tr. Accessed:July 12, 2026
[16]. Wang, J., Zeng, G., Xu, M., Wan, X., Wang, K., Mou, J., Hua, C., Fan, C., Han, P. 2024.SSA-
[17]. BP Neural Network Model for Predicting Rice-Fish Production in China. Journal of Applied Ichthyology, 2024(1), 5739961. https://doi.org/10.1155/2024/5739961
[18]. White PJ, Xing Y. 1997. Antioxidants from Cereals and Legumes. Editor: Shahidi F. Natural
[19]. Antioxidants: Chemistry, Health Effects, and Applications, Champaign IL: AOCS Press
[20]. Wood M. 2001. New Oats and Barleys, Ready for Breakfast, Brewery or Bran, Agricultural Research, 49(8), 18-19.
[21]. Zhang, W., Bai, C., Liu, G. 2007. A longer-term forecast on global supply and demand of food products. Journal of Food Agriculture and Environment, 5(1), 105-110.