Correlation analysis between the Gross Domestic Product and the production of dairy products in Colombia between 2005 and 2022

William Niebles N[1],José Marcelo Torres Ortega[2],Claudia Rojas Martínez[3]

1.*Universidad de Sucre,williamniebles@yahoo.com.mx*
2.*Universidad de Sucre*
3.*Universidad de Sucre*

ABSTRACT

The purpose of this study is to examine the connection between Colombia's GDP and dairy product production from 2005 to 2022. R Studio was used to process the data, which came from the Republic of Colombia's National Administrative Department of Statistics (DANE). In order to employ the Pearson correlation coefficient in the second stage of the study, the Shapiro-Wilk test was first used to confirm that the data was normal. At the interval or ratio level, this coefficient quantifies the relationship between two variables. According to the findings, GDP and the production of dairy products have a substantial positive association (r = 0.88), and both variables are regularly distributed. Additionally, the association is statistically significant, as indicated by the hypothesis test's p-value of less than 0.05. The study suggests that as GDP increases or decreases, dairy production follows a similar trend, which could be related to economic demand in different contexts. However, it is noted that this correlation does not imply direct causality, since other factors could influence milk production.
Key words: Gross Domestic Product, Dairy products, Correlation, Statistics, Economic growth.

INTRODUCTION

Dairy production has long been a key sector in the global economy, not only for its contribution to food supply, but also for its impact on industrial and agricultural development. Dairy products, which include a wide range of milk-derived foods such as cheese, butter, yogurt and processed milk, represent an important source of nutrients in the global diet, while their production and distribution contribute considerably to national income (Stobiecka et al., 2022). In many countries, dairy production has proven to be a key factor in boosting employment, especially in rural areas, where much of the milk production takes place. This links dairy production not only to food systems, but also to the economic and social development of many regions (Arvidsson et al., 2020). In this sense, changes in a country's Gross Domestic Product (GDP), as a general measure of economic growth, can have a significant relationship with the quantity of dairy products produced, since a growing GDP tends to reflect greater consumption and production of goods (Irtyshcheva et al., 2020).

At a global level, fluctuations in the economy directly affect dairy production, since the increase or decrease in the purchasing power of the population can influence the consumption of these products. In times of economic expansion, there is an increase in the demand for higher value-added products, including dairy products, since consumers have greater purchasing power to opt for more expensive or diversified foods (Dahal et al., 2024). On the other hand, during periods of economic recession, the demand for these products may decrease, affecting the profitability of the dairy industry. This relationship is especially evident in developed countries, where the consumption of dairy products is deeply rooted in the diet and is influenced by changes in disposable income (Ecker and Pauw 2024).

William Niebles N[1], José Marcelo Torres Ortega[2], Claudia Rojas Martínez[3]

In this global context, some studies have suggested a correlation between GDP and dairy production, suggesting that economic growth tends to be linked to an increase in the production and consumption of dairy products. However, the particularities of each country and its economic structure influence the nature and strength of this relationship. Domestic demand for dairy products can be influenced by several factors, such as the level of industrialization, government policies, trade barriers, access to international markets, and the dynamics of the national market (Anderson 2022).

In the case of Colombia, the dairy sector is of strategic importance within the national economy. Milk production and the processing of dairy products are fundamental activities for the country's agricultural and agro-industrial sector, especially in rural areas (Beneberu and Wondifraw 2020). Colombia has a vast rural population dedicated to milk production, and the country has developed a considerable industrial infrastructure to process these products. This situation makes the dairy industry not only contribute to the national GDP, but also essential for the livelihood of thousands of rural families and for regional economic development (Khan et al., 2020).

Historically, dairy production in Colombia has been subject to fluctuations derived from various economic, social and political factors. For example, trade liberalization in the 1990s affected the local market, exposing it to foreign competition, which, in many cases, has negatively impacted local dairy production (Liu 2024). Despite these challenges, the sector has managed to recover and remains an essential component of the Colombian economy. Furthermore, government policies to strengthen agricultural production, along with incentives for small and medium-sized producers, have helped boost the production of dairy products in recent years (Monroy-Gomez et al., 2022).

The relationship between GDP and dairy production in Colombia can be complex, given that the Colombian economy has gone through different phases of growth and recession over the past two decades. Between 2005 and 2022, the country experienced periods of sustained economic growth, particularly driven by the natural resource bonanza and foreign investments (Nahrin et al., 2023). During these years, Colombian GDP grew significantly, which could have contributed to an increase in dairy production and consumption. However, it is also important to consider that Colombia has faced significant economic challenges, such as the 2008 global financial crisis and the economic impact of the COVID-19 pandemic in 2020 and 2021, which affected both GDP growth and the production of goods in general (Ramírez et al., 2023).

In this context, it is relevant to explore how dairy production has responded to GDP variations in Colombia during the period between 2005 and 2022. The hypothesis put forward in this study suggests that GDP growth or decrease may be related to the increase or decrease in dairy production, given that an expanding economy generally favors greater food production and consumption. However, since the dairy sector is also influenced by factors such as domestic demand, government policies, and production costs, it is essential to carry out a more detailed analysis to confirm whether there is a significant relationship between these two variables.

To determine whether there is a significant relationship between Colombia's GDP and dairy production from 2005 to 2022, this article conducts a correlation study. Statistics such as the Pearson correlation test were used to determine the strength and direction of the link between the two variables using secondary data (Haddad et al., 2021), provided by DANE. This strategy will not only help us understand the relationship between Colombia's economic development and dairy production, but will also serve as a basis for further studies examining other relevant aspects of this industry.

The analysis also seeks to provide useful information for policy makers and actors in the agro-industrial sector, in order to understand how the country's macroeconomic behavior influences specific sectors, such as the dairy industry. Likewise, the results of this study are

expected to offer a broader view of how changes in GDP can impact the dairy industry, which would allow for the design of production and marketing strategies better adapted to economic cycles. Through this correlational analysis, relevant conclusions for the sector and recommendations can be offered to improve the competitiveness and sustainability of the dairy industry in Colombia.

MATERIALS AND METHODS

The objective of this analysis is to determine if there is a significant association between the variables "Gross Domestic Product (GDP)" and "Manufacture of dairy products in Colombia from 2005 to 2022". The data used were obtained from secondary information sources and the two variables were taken from the DANE website (2024). The first variable represents GDP growth, measured in millions of euros per year, and the second comprises the amount of dairy products produced in Colombia between 2005 and 2022. The data processing was carried out through the statistical software R Studio Team (2012).

First, as the first methodological phase, the Shapiro-Wilk test was applied to the two variables mentioned above. This test was chosen because the sample size is less than 50 data (de Souza et al., 2023).

Second, because the variables were regularly distributed, the Pearson correlation analysis was performed. This statistical test examines how two variables are related, measured at the interval or ratio level. The "product-moment coefficient" is another name for it (Ma and Qin, 2020). It is calculated using the scores of the two variables in a sample. In the same individuals or situations, the scores obtained from one variable are connected with the scores obtained from the other (Sürücü and Maslakci, 2020).

In correlational research based on approaches and hypotheses that link variables, correlation tests are used. The research hypothesis postulates that there is a significant relationship between the variables and the null hypothesis that there is not (Komaroff 2020). The selection of the appropriate correlation test or coefficient depends on the measurement level of the variables to be associated; when the variables are interval or ratio, this coefficient is frequently used (Baak 2020).

Regarding correlation coefficients, these can vary from −1.00 to 1.00. For any coefficient between −1.00 and 0, Y always decreases by a fixed amount when X increases by one unit; this is equally true for both "the smaller the X, the greater the Y" and for "the larger the X, the smaller the Y" proportionally. Meanwhile, for each coefficient in the range of 0 to 1.00, "the larger the X, the greater the Y" or "the smaller the X, the smaller the Y". In other words, Y always increases by the same amount whenever X increases, just as it does when X decreases. Although these interpretations vary in their writing from one author to another, they are all consistent as to the meaning of each possible outcome (Hernández-Sampieri and Mendoza, 2018).

Depending on each value of the correlation coefficient, also according to Hernández-Sampieri and Mendoza (2018), the level of correlation is classified according to the following scale: "−1.00" = "Perfect negative" correlation; "−0.90" = "Very strong negative" correlation; "−0.75" = "Considerable negative" correlation; "−0.50" = "Medium negative" correlation; "−0.25" = "Weak negative" correlation; "−0.10" = "Very weak negative" correlation; "0.00" = "There is no correlation between the variables"; "0.10" = "Very weak positive" correlation; "0.25" = "Weak positive" correlation; "0.50" = "Medium positive" correlation; "0.75" = "Considerable positive" correlation; "0.90" = "Very strong positive" correlation, and finally "1.00" = "Perfect positive" correlation.

Finally, as a third methodological phase, to determine whether the observed correlation was statistically significant, a hypothesis test was performed. The null hypothesis (H_0) established that the variables were not correlated with each other, while the alternative hypothesis (H_1) stated that a significant correlation was evident (Ramírez et al., 2022).

William Niebles N[1],José Marcelo Torres Ortega[2],Claudia Rojas Martínez[3]

RESULTS
Table 1

Data of the dependent variable (EPL) and the independent variable (GDP) from 2005 to 2022

EPL	PIB
2609	514853
2659	549435
2726	586457
2880	605713
2855	612616
2838	640151
2947	684628
2981	711415
3139	747939
3177	781589
3217	804692
3271	821489
3301	832656
3312	854008
3371	881224
3405	817900
3127	906243
3159	972298

Source: DANE (2024)

Normality test

```
> shapiro.test(datos$EPL)

        Shapiro-Wilk normality test

data:  datos$EPL
W = 0.942, p-value = 0.3135

> shapiro.test(datos$PIB)

        Shapiro-Wilk normality test

data:  datos$PIB
W = 0.96355, p-value = 0.6714
```

Figure 1. *Results of the Shapiro-Wilk normality test in R Studio. Source: authors using R Studio (2024).*

Since both variables under study have a P value greater than 0.05, Figure 1 shows the results of the normality test applied to the variables "Gross Domestic Product" and "Manufacture of dairy products in Colombia from 2005 to 2022". Consequently, it can be said that their distribution is normal.

Pearson correlation coefficient

As can be seen in Figure 2, the hypothesis test yielded a P value of 0.000001075. Since the P value is less than the significance level of $\alpha = 0.05$, the null hypothesis is rejected in favor of the alternative, i.e., the variables are correlated. This also suggests that the observed association is statistically significant and not a product of chance. The Pearson coefficient calculated to estimate the correlation between the variables "Gross Domestic Product" and "Manufacture of dairy products in Colombia from 2005 to 2022" was 0.88. This value
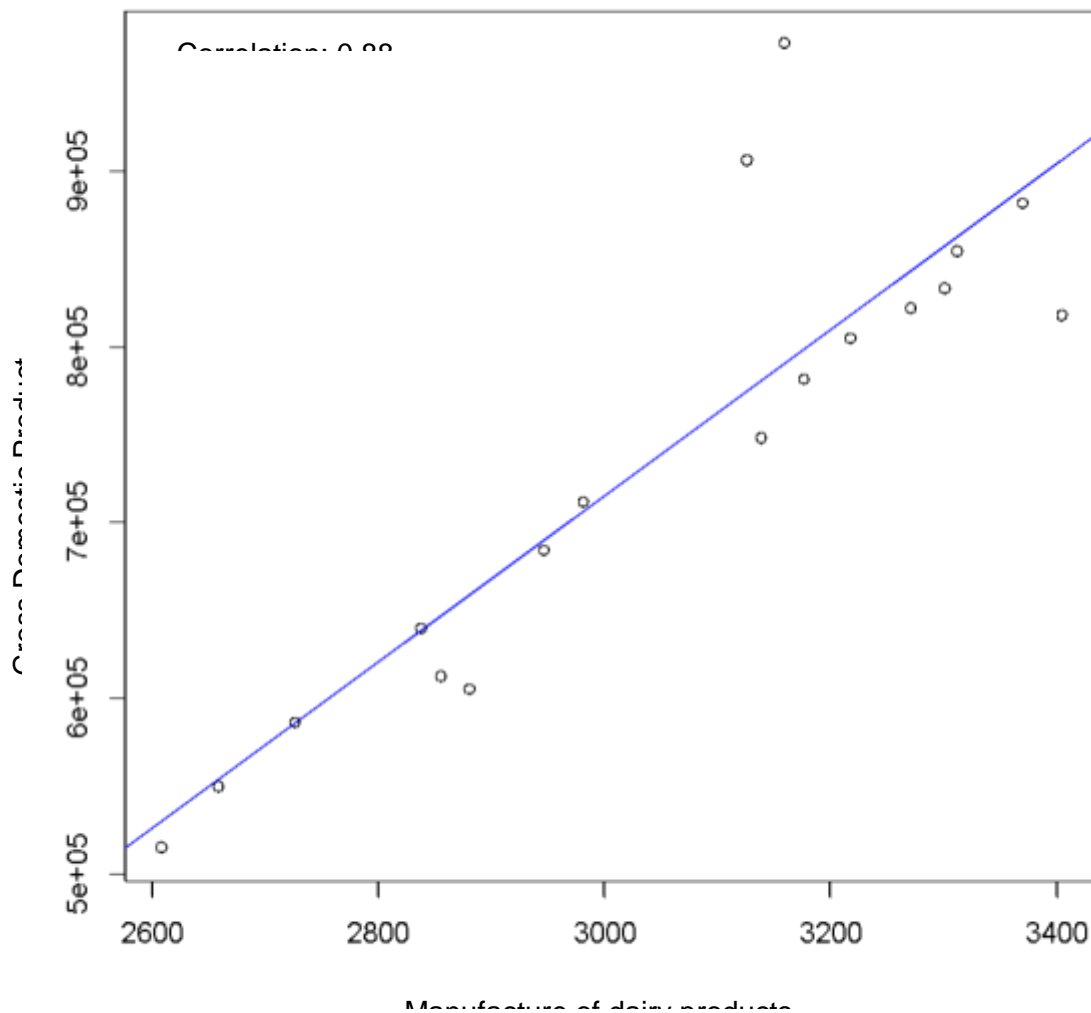
319

suggests a strong positive correlation; the linearity and value of the correlation can be reaffirmed by observing Figure 3.

The results indicate that, in general, as Colombia's GDP between 2005 and 2022 increases or decreases, there has been an increase or decrease in the amount of dairy products produced, respectively. This could suggest that, in periods of economic growth, more dairy products are produced. On the other hand, in times of economic recession, dairy production may decrease, perhaps due to low demand for the products.

```
          Pearson's product-moment correlation

data:  datos$EPL and datos$PIB
t = 7.5981, df = 16, p-value = 1.075e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7122399 0.9565647
sample estimates:
       cor
0.8848699
```

Figure 2. *Results of the Pearson test performed in R Studio. Source: authors using R Studio (2024).*

Figure 3. *Scatter plot made in R Studio. Source: authors using R Studio (2024).*

320

William Niebles N[1],José Marcelo Torres Ortega[2],Claudia Rojas Martínez[3]

As can be seen in Figure 3, the level of correlation obtained in the previous paragraphs between the variables studied is confirmed. It is advisable to understand that the relationship between the variables does not imply causality. Although there is a strong association between them, it cannot be inferred that GDP growth is the direct cause of the increase or decrease in the quantity of dairy products produced. Other factors of different kinds could be influencing this relationship in one way or another.

CONCLUSIONS

This study examined the relationship between Colombia's GDP and dairy production from 2005 to 2022 in order to determine whether or not there is a substantial correlation between the two variables. This was accomplished using DANE data and a variety of statistical tests, yielding significant findings that helped to understand the relationship between the country's dairy production and economic growth.

First, the results of the Shapiro-Wilk normality test indicated that both GDP and dairy production are normally distributed. This finding allowed the application of Pearson's correlation, ensuring the validity of statistical inferences in this context. This methodological step was key to grounding the correlational analysis.

A strong positive connection between the variables is indicated by the Pearson correlation coefficient result of 0.88. This implies that, generally speaking, dairy production increases in tandem with Colombia's GDP and vice versa. This result supports the assumption that economic growth leads to an increase in the production of goods (in this case, dairy products). Finally, since the p-value was significantly below the significance level $\alpha=0.05$, the hypothesis test verified that the observed association was not the result of chance. This result reaffirms the importance of the data examined and supports the conclusion that there is a statistically significant association between GDP and dairy production.

In conclusion, the study shows a robust association between economic growth and dairy production in Colombia. However, it is important to note that this correlation does not imply direct causality. Although the results show a strong relationship between the variables, other factors, such as changes in demand, government policies or fluctuations in input prices, could influence dairy production. Therefore, future studies should further explore these additional variables to comprehensively understand the phenomenon.

This analysis represents a valuable contribution to the understanding of the dairy industry in Colombia, suggesting that economic growth may be a relevant indicator to predict trends in food production. However, the complexity of the economic and social context implies that a direct linear relationship should not be assumed, but rather multiple factors should be considered in future research.

REFERENCES

Allaire, J. (2012). RStudio: integrated development environment for R. *Boston, MA, 770*(394), 165-171. https://n9.cl/5zn1y

Anderson, K. (2022). Trade-related food policies in a more volatile climate and trade environment. *Food Policy, 109*, 102253. https://doi.org/10.1016/j.foodpol.2022.102253

Arvidsson Segerkvist, K., Hansson, H., Sonesson, U., & Gunnarsson, S. (2020). Research on environmental, economic, and social sustainability in dairy farming: A systematic mapping of current literature. *Sustainability, 12*(14), 5502. https://doi.org/10.3390/su12145502

Baak, M., Koopman, R., Snoek, H., & Klous, S. (2020). A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. *Computational Statistics & Data Analysis, 152*, 107043. https://doi.org/10.1016/j.csda.2020.107043

Beneberu, N., & Wondifraw, Z. (2020). Challenges and opportunities of milk production in urban dairy farms in ethiopia: a review. Food Sci. Qual. *Manage, 102,* 12-20. https://doi.org/10.7176/FSQM/102-02

Dahal, B., Kimmerer, C., & Hailu, G. (2024). Value added to agricultural commodities. In *Future Food Systems* (pp. 89-107). Academic Press https://doi.org/10.1016/B978-0-443-15690-8.00010-2

Departamento Administrativo Nacional de Estadística (DANE) (15 de agosto de 2024). *Producto Interno Bruto (PIB) nacional trimestral*. Recuperado el 18 de septiembre de 2024 de: https://n9.cl/xp3x

de Souza, R. R., Toebe, M., Mello, A. C., & Bittencourt, K. C. (2023). Sample size and Shapiro-Wilk test: An analysis for soybean grain yield. *European Journal of Agronomy, 142*, 126666. https://doi.org/10.1016/j.eja.2022.126666

Ecker, O., & Pauw, K. (2024). Dairy consumption and household diet quality in East Africa: Evidence from survey-based simulation models. *Food Policy, 122*, 102562. https://doi.org/10.1016/j.foodpol.2023.102562

Haddad, M., Sanders, D., Tewkesbury, G., & Bausch, N. (2021). Integrating statistical correlation with discrete multi-criteria decision-making. *International Journal of Information and Decision Sciences, 13*(1), 1-15. https://doi.org/10.1504/IJIDS.2021.113599

Hernández-Sampieri, R., & Mendoza, C. (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*. México: McGraw Hill. https://n9.cl/tw9bs

Irtyshcheva, I., Stehnei, M., Popadynets, N., Danylo, S., Rogatina, L., Bogatyrev, K., ... & Voit, O. (2020). Business process management in the food industry under the conditions of economic transformations. *Management Science Letters, 10*(14), 3243-3252. http://dx.doi.org/10.5267/j.msl.2020.6.014

Khan, N., Fahad, S., Naushad, M., & Faisal, S. (2020). COVID-2019 locked down impact on dairy industry in the world. *Available at SSRN* 3616325. https://dx.doi.org/10.2139/ssrn.3616325

Komaroff, E. (2020). Relationships between p-values and Pearson correlation coefficients, type 1 errors and effect size errors, under a true null hypothesis. J*ournal of Statistical Theory and Practice, 14*(3), 49. https://doi.org/10.1007/s42519-020-00115-6

Liu, X., Liang, Y., & Chen, K. Z. (2024). Dairy trade liberalization and child stunting: Evidence from low-and middle-income countries. *Food Policy, 122*, 102554. https://doi.org/10.1016/j.foodpol.2023.102554

Ma, X., & Qin, H. (2020). A new parameter reduction algorithm for interval-valued fuzzy soft sets based on Pearson's product moment coefficient. *Applied Intelligence, 50*(11), 3718-3730. https://doi.org/10.1007/s10489-020-01708-1

Monroy-Gomez, J., Ferraboschi, C., van Zutphen, K. G., Gavin-Smith, B., Amanquah, D., & Kraemer, K. (2022). Small and medium enterprises' perspectives on food fortification amid the growing burden of malnutrition. *Nutrients, 14*(18), 3837. https://doi.org/10.3390/nu14183837

Nahrin, R., Rahman, M. H., Majumder, S. C., & Esquivias, M. A. (2023). Economic growth and pollution nexus in Mexico, Colombia, and Venezuela (G-3 countries): the role of renewable energy in carbon dioxide emissions. *Energies, 16*(3), 1076. https://doi.org/10.3390/en16031076

Ramírez, J., Ballestas, M., Herrera, H., & Ballesta, I. (2023). Pobreza multidimensional y pobreza monetaria de los migrantes venezolanos: Un estudio en Barranquilla, Colombia. *Revista de ciencias sociales, 29*(2), 373-386. https://dialnet.unirioja.es/servlet/articulo?codigo=8920557

Ramírez Duran, J., Ballestas, M., Henry, H., & Ballesta, I. (2022). Factores asociados al desempleo de los migrantes venezolanos en la ciudad de Barranquilla, Colombia. *Revista de Ciencias Sociales (RCS), 28*(2), 150-165. https://doi.org/10.31876/rcs.v28i2.37926

Stobiecka, M., Król, J., & Brodziak, A. (2022). Antioxidant activity of milk and dairy products. *Animals, 12*(3), 245. https://doi.org/10.3390/ani12030245

Sürücü, L., & Maslakci, A. (2020). Validity and reliability in quantitative research. Business & Management Studies: *An International Journal, 8*(3), 2694-2726. https://doi.org/10.15295/bmij.v8i3.1540